



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>7</sup> :</b> <b>C12N 15/12, C07K 14/47, C12Q 1/68,</b> <b>C07K 16/18, A61K 38/17</b>	<b>A2</b>	<b>(11) International Publication Number:</b> <b>WO 00/17355</b> <b>(43) International Publication Date:</b> 30 March 2000 (30.03.00)
<b>(21) International Application Number:</b> PCT/US99/21565 <b>(22) International Filing Date:</b> 17 September 1999 (17.09.99)  <b>(30) Priority Data:</b> 60/172,226                      18 September 1998 (18.09.98)      US 60/131,321                      27 April 1999 (27.04.99)              US  <b>(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications</b> US    60/172,226 (CIP) Filed on                                      18 September 1998 (18.09.98) US    60/131,321 (CIP) Filed on                                      27 April 1999 (27.04.99)  <b>(71) Applicant (for all designated States except US):</b> INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).  <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> LAL, Preeti [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US). TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). HILLMAN, Jennifer, L. [US/US]; 230		Monroe Drive #12, Mountain View, CA 94040 (US). BANDMAN, Olga [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). CORLEY, Neil, C. [US/US]; 1240 Dale Avenue #30, Mountain View, CA 94040 (US). GUEGLER, Karl, J. [CH/US]; 1048 Oakland Avenue, Menlo Park, CA 94025 (US). PATTERSON, Chandra [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). AZIMZAI, Yalda [US/US]; 2045 Rock Springs Drive, Hayward, CA 94545 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US).  <b>(74) Agents:</b> BILLINGS, Lucy, J. et al.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).  <b>(81) Designated States:</b> AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).  <b>Published</b> <i>Without international search report and to be republished upon receipt of that report.</i>
<b>(54) Title:</b> HUMAN CYTOSKELETON ASSOCIATED PROTEINS  <b>(57) Abstract</b> <p>The invention provides human cytoskeleton associated proteins (CYSKP) and polynucleotides which identify and encode CYSKP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of CYSKP.</p>		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

<b>AL</b>	Albania	<b>ES</b>	Spain	<b>LS</b>	Lesotho	<b>SI</b>	Slovenia
<b>AM</b>	Armenia	<b>FI</b>	Finland	<b>LT</b>	Lithuania	<b>SK</b>	Slovakia
<b>AT</b>	Austria	<b>FR</b>	France	<b>LU</b>	Luxembourg	<b>SN</b>	Senegal
<b>AU</b>	Australia	<b>GA</b>	Gabon	<b>LV</b>	Latvia	<b>SZ</b>	Swaziland
<b>AZ</b>	Azerbaijan	<b>GB</b>	United Kingdom	<b>MC</b>	Monaco	<b>TD</b>	Chad
<b>BA</b>	Bosnia and Herzegovina	<b>GE</b>	Georgia	<b>MD</b>	Republic of Moldova	<b>TG</b>	Togo
<b>BB</b>	Barbados	<b>GH</b>	Ghana	<b>MG</b>	Madagascar	<b>TJ</b>	Tajikistan
<b>BE</b>	Belgium	<b>GN</b>	Guinea	<b>MK</b>	The former Yugoslav Republic of Macedonia	<b>TM</b>	Turkmenistan
<b>BF</b>	Burkina Faso	<b>GR</b>	Greece	<b>ML</b>	Mali	<b>TR</b>	Turkey
<b>BG</b>	Bulgaria	<b>HU</b>	Hungary	<b>MN</b>	Mongolia	<b>TT</b>	Trinidad and Tobago
<b>BJ</b>	Benin	<b>IE</b>	Ireland	<b>MR</b>	Mauritania	<b>UA</b>	Ukraine
<b>BR</b>	Brazil	<b>IL</b>	Israel	<b>MW</b>	Malawi	<b>UG</b>	Uganda
<b>BY</b>	Belarus	<b>IS</b>	Iceland	<b>MX</b>	Mexico	<b>US</b>	United States of America
<b>CA</b>	Canada	<b>IT</b>	Italy	<b>NE</b>	Niger	<b>UZ</b>	Uzbekistan
<b>CF</b>	Central African Republic	<b>JP</b>	Japan	<b>NL</b>	Netherlands	<b>VN</b>	Viet Nam
<b>CG</b>	Congo	<b>KE</b>	Kenya	<b>NO</b>	Norway	<b>YU</b>	Yugoslavia
<b>CH</b>	Switzerland	<b>KG</b>	Kyrgyzstan	<b>NZ</b>	New Zealand	<b>ZW</b>	Zimbabwe
<b>CI</b>	Côte d'Ivoire	<b>KP</b>	Democratic People's Republic of Korea	<b>PL</b>	Poland		
<b>CM</b>	Cameroon	<b>KR</b>	Republic of Korea	<b>PT</b>	Portugal		
<b>CN</b>	China	<b>KZ</b>	Kazakstan	<b>RO</b>	Romania		
<b>CU</b>	Cuba	<b>LC</b>	Saint Lucia	<b>RU</b>	Russian Federation		
<b>CZ</b>	Czech Republic	<b>LI</b>	Liechtenstein	<b>SD</b>	Sudan		
<b>DE</b>	Germany	<b>LK</b>	Sri Lanka	<b>SE</b>	Sweden		
<b>DK</b>	Denmark	<b>LR</b>	Liberia	<b>SG</b>	Singapore		
<b>EE</b>	Estonia						

## HUMAN CYTOSKELETON ASSOCIATED PROTEINS

### TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of human cytoskeleton  
5 associated proteins and to the use of these sequences in the diagnosis, treatment, and prevention of  
cell proliferative, autoimmune/inflammatory, vesicle trafficking, neurological, cell motility,  
reproductive, and muscle disorders.

### BACKGROUND OF THE INVENTION

10 The cytoskeleton, a cytoplasmic system of protein fibers, mediates cell shape, structure, and  
movement. The cytoskeleton supports the cell membrane and forms tracks along which organelles  
and other elements move in the cytosol. The cytoskeleton is a dynamic structure that allows cells to  
adopt various shapes and to carry out directed movements. Major cytoskeletal fibers are the  
microfilaments, the microtubules, and the intermediate filaments. Motor proteins, including myosin,  
15 dynein, and kinesin, drive movement of, or along, the fibers. The motor protein dynamin drives the  
formation of membrane vesicles. Accessory or associated proteins modify the structure or activity of  
the fibers while cytoskeletal membrane anchors connect the fibers to the cell membrane. (The  
cytoskeleton is reviewed in Lodish, H. et al. (1995) Molecular Cell Biology Scientific American  
Books, New York NY.)

20

#### Microtubules and Associated Proteins

##### Tubulins

Microtubules, cytoskeletal fibers with a diameter of 24 nm, have multiple roles in the cell.  
Bundles of microtubules form cilia and flagella, which are whip-like extensions of the cell membrane  
25 that are necessary for sweeping materials across an epithelium and for swimming of sperm,  
respectively. Marginal bands of microtubules in red blood cells and platelets are important for these  
cells' pliability. Organelles, membrane vesicles, and proteins are transported in the cell along tracks  
of microtubules. For example, microtubules run through nerve cell axons, allowing bi-directional  
transport of materials and membrane vesicles between the cell body and the nerve terminal. Failure  
30 to supply the nerve terminal with these vesicles blocks the transmission of neural signals.  
Microtubules, in the form of the spindle, are also critical to chromosomal movement during cell  
division. Both stable and short-lived populations of microtubules exist in the cell.

Microtubules are a polymer of GTP-binding tubulin protein subunits. Each subunit is a  
heterodimer of  $\alpha$ - and  $\beta$ - tubulin, multiple isoforms of which exist. The hydrolysis of GTP is linked

to the addition of tubulin subunits at the end of a microtubule. The subunits interact head to tail to form protofilaments; the protofilaments interact side to side to form a microtubule. A microtubule is polarized, one end ringed with  $\alpha$ -tubulin and the other with  $\beta$ -tubulin, and the two ends differ in their rates of assembly. Generally each microtubule is composed of 13 protofilaments although 11 or 15 protofilament-microtubules are sometimes found. Cilia and flagella contain doublet microtubules. Microtubules grow from specialized structures known as centrosomes or microtubule-organizing centers (MTOCs). MTOCs may contain one or two centrioles, which are pinwheel arrays of triplet microtubules. The basal body, the organizing center located at the base of a cilium or flagellum, contains one centriole.  $\gamma$ -tubulin present in the MTOC is important for nucleating the polymerization of  $\alpha$ - and  $\beta$ -tubulin heterodimers but does not polymerize into microtubules. The protein pericentrin is found in the MTOC and has a role in microtubule assembly.

#### Microtubule-Associated Proteins

Microtubule-associated proteins (MAPs) have roles in the assembly and stabilization of microtubules. One major family of MAPs, assembly MAPs, can be identified in neurons as well as non-neuronal cells. Assembly MAPs are responsible for cross-linking microtubules in the cytosol. These MAPs are organized into two domains: a basic microtubule-binding domain and an acidic projection domain. The projection domain is the binding site for membranes, intermediate filaments, or other microtubules. Based on sequence analysis, assembly MAPs can be further grouped into two types: Type I and Type II.

Type I MAPs, which include MAP1A and MAP1B, are large, filamentous molecules that copurify with microtubules and are abundantly expressed in brain and testis. They contain several repeats of a positively-charged amino acid sequence motif that binds and neutralizes negatively charged tubulin, leading to stabilization of microtubules. MAP1A and MAP1B are each derived from a single precursor polypeptide that is subsequently proteolytically processed to generate one heavy chain and one light chain.

Another light chain, LC3, is a 16.4 kDa molecule that binds MAP1A, MAP1B, and microtubules. It is suggested that LC3 is synthesized from a source other than the MAP1A or MAP1B transcripts, and the expression of LC3 may be important in regulating the microtubule binding activity of MAP1A and MAP1B during cell proliferation (Mann, S. S. et al. (1994) J. Biol. Chem. 269:11492-11497).

Type II MAPs, which include MAP2a, MAP2b, MAP2c, MAP4, and Tau, are characterized by three to four copies of an 18-residue sequence in the microtubule-binding domain. MAP2a, MAP2b, and MAP2c are found only in dendrites, MAP4 is found in non-neuronal cells, and Tau is found in axons and dendrites of nerve cells. Alternative splicing of the Tau mRNA leads to the

existence of multiple forms of Tau protein. Tau phosphorylation is altered in neurodegenerative disorders such as Alzheimer's disease, Pick's disease, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia and Parkinsonism linked to chromosome 17. The altered Tau phosphorylation leads to a collapse of the microtubule network and the formation of

5 intraneuronal Tau aggregates (Spillantini, M.G. and Goedert, M. (1998) Trends Neurosci. 21:428-433).

Tektins are filamentous proteins that were originally discovered in association with axonemal microtubules of sea urchin sperm. Subsequent work has shown that tektins are also found in association with spindle microtubules in clams and in mammals. (Steffen, W. and Linck, R.W. (1992) J. Cell Sci. 101:809-822.) Tektins may form rod-like alpha-helical structures similar to those of intermediate filament proteins (Norrander, J.M. et al. (1996) J. Mol. Biol. 29:385-397).

Microtubular aggregates are associated with several disorders. An extraskeletal myxoid chondrosarcoma tumor from human contained parallel arrays of microtubules within the rough endoplasmic reticulum (Suzuki, T. et al. (1988) J. Pathol. 156:51-57). Microtubular aggregates were

15 also found in hepatocytes from chimpanzees infected with hepatitis C virus. Monoclonal antibodies prepared to these aggregates detect a protein called p44 (or microtubular aggregates protein) (Maeda, T. et al. (1989) J. Gen. Virol. 70:1401-1407). A human homolog of p44 is inducible by interferon- $\alpha$  and interferon- $\beta$ , but not by interferon- $\gamma$ . p44 protein may be a mediator in the antiviral action of interferon (Kitamura, A. et al. (1994) Eur. J. Biochem. 224:877-883).

## 20 Dynein-related Motor Proteins

Dyneins are (-) end-directed motor proteins which act on microtubules. Two classes of dyneins exist, cytosolic and axonemal. Cytosolic dyneins are responsible for translocation of materials along cytoplasmic microtubules, for example, transport from the nerve terminal to the cell body and transport of endocytic vesicles to lysosomes. Cytoplasmic dyneins are also reported to play

25 a role in mitosis. Axonemal dyneins are responsible for the beating of flagella and cilia. Dynein on one microtubule doublet walks along the adjacent microtubule doublet. This sliding force produces bending forces that cause the flagellum or cilium to beat. Dyneins have a native mass between 1000 and 2000 kDa and contain either two or three force-producing heads driven by the hydrolysis of ATP. The heads are linked via stalks to a basal domain which is composed of a highly variable number of

30 accessory intermediate and light chains.

## Microfilaments and Associated Proteins

### Actins

Microfilaments, cytoskeletal filaments with a diameter of 7-9 nm, are vital to cell locomotion.

cell shape, cell adhesion, cell division, and muscle contraction. Assembly and disassembly of the microfilaments allow cells to change their morphology. Microfilaments are the polymerized form of actin, the most abundant intracellular protein in the eukaryotic cell. Human cells contain six isoforms of actin. The three  $\alpha$ -actins are found in different kinds of muscle, nonmuscle  $\beta$ -actin and nonmuscle  $\gamma$ -actin are found in nonmuscle cells, and another  $\gamma$ -actin is found in intestinal smooth muscle cells. G-actin, the monomeric form of actin, polymerizes into polarized, helical F-actin filaments, accompanied by the hydrolysis of ATP to ADP. Actin filaments associate to form bundles and networks, providing a framework to support the plasma membrane and determine cell shape. These bundles and networks are connected to the cell membrane. In muscle cells, thin filaments containing actin slide past thick filaments containing the motor protein myosin during contraction. A family of actin-related proteins exist that are not part of the actin cytoskeleton, but rather associate with microtubules and dynein.

#### Actin-Associated Proteins

Actin-associated proteins have roles in cross-linking, severing, and stabilization of actin filaments and in sequestering actin monomers. Several of the actin-associated proteins have multiple functions. Bundles and networks of actin filaments are held together by actin cross-linking proteins. These proteins have two actin-binding sites, one for each filament. Short cross-linking proteins promote bundle formation while longer, more flexible cross-linking proteins promote network formation. Calmodulin-like calcium-binding domains in actin cross-linking proteins allow calcium regulation of cross-linking. Group I cross-linking proteins have unique actin-binding domains and include the 30 Kd protein, EF-1a, fascin, and scruin. Group II cross-linking proteins have a 7,000-MW actin-binding domain and include villin and dematin. Group III cross-linking proteins have pairs of a 26,000-MW actin-binding domain and include fimbrin, spectrin, dystrophin, ABP 120, and filamin.

Severing proteins regulate the length of actin filaments by breaking them into short pieces or by blocking their ends. Severing proteins include gCAP39, severin (fragmin), gelsolin, and villin. Capping proteins can cap the ends of actin filaments, but cannot break filaments. Capping proteins include CapZ, tropomodulin, and tensin.

Tensin, which is found in focal adhesions, also crosslinks actin filaments. Integrin activation by the extracellular matrix leads to the phosphorylation of tensin on tyrosine, serine, and threonine residues; this phosphorylation also occurs in cells transformed with oncogenes. Tensin has an SH2 domain and may bind to other tyrosine-phosphorylated proteins. (Lo, S.H. et al. (1997) J. Cell Biol. 136:1349-1361.) The N-terminus of tensin contains a region homologous to the catalytic domain of a putative tyrosine phosphatase (PTP) from Saccharomyces cerevisiae. This PTP domain in tensin may

mediate binding interactions with phosphorylated polypeptides (Haynie, D.T. and Ponting, C.P. (1996) Protein Sci. 5:2643-2646). Mice which lack the tensin gene have kidney abnormalities, indicating that the loss of tensin leads to weakening of focal adhesions in the kidney (Lo, supra).

The proteins thymosin and profilin sequester actin monomers in the cytosol, allowing a pool  
 5 of unpolymerized actin to exist. Profilin may also stimulate F-actin formation by effectively lowering the critical concentration required for actin monomer addition (Gertler, F.B. et al. (1996) Cell 87:227-239).

The Ena/VASP (vasodilator-stimulated phosphoprotein) protein family has roles in actin-based motility. These proteins, including Mena, VASP, and Evl (Ena/VASP-like), have homology to  
 10 the Drosophila Enabled protein which is involved in neural development. Mammalian Ena/VASP proteins localize at focal contacts and in regions where actin filaments are highly dynamic. The neural forms of Mena induce F-actin rich outgrowths in fibroblasts. Mena may have roles in microfilament-based extension of filopodia during axonal growth cone migration. In vitro motility assays of the intracellular pathogenic bacterium Listeria monocytogenes in platelet and brain extracts  
 15 show that the Ena/VASP proteins play interchangeable roles in the transformation of actin polymerization into active movement and propulsive force. The Ena/VASP proteins associate with actin, profilin, the focal adhesion protein zyxin, and vinculin. Phosphorylation of Mena and VASP may regulate their activity. (Gertler, supra; Laurent, V. et al. (1999) J. Cell Biol. 144:1245-1258.)

The actin-associated proteins tropomyosin, troponin, and caldesmon regulate muscle  
 20 contraction in response to calcium. The tropomyosin proteins, found in muscle and nonmuscle cells, are  $\alpha$ -helical and form coiled-coil dimers. Striated muscle tropomyosin mediates the interactions between the troponin complex and actin, regulating muscle contraction. (PROSITE PDOC00290 Tropomyosins signature.) The troponin complex is composed of troponin-T, troponin-I, and troponin-C. Troponin-T binds tropomyosin, linking troponin-I and troponin-C to tropomyosin.

25

#### Intermediate Filaments and Associated Proteins

Intermediate filaments (IFs) are cytoskeletal fibers with a diameter of 10 nm, intermediate between that of microfilaments and microtubules. They serve structural roles in the cell, reinforcing cells and organizing cells into tissues. IFs are particularly abundant in epidermal cells and in neurons.  
 30 IFs are extremely stable, and, in contrast to microfilaments and microtubules, do not function in cell motility. IF proteins include acidic keratins, basic keratins, desmin, glial fibrillary acidic protein, vimentin, peripherin, neurofilaments, nestin, and lamins.

IFs have a central  $\alpha$ -helical rod region interrupted by short nonhelical linker segments. The rod region is bracketed, in most cases, by non-helical head and tail domains. The rod regions of

intermediate filament proteins associate to form a coiled-coil dimer. A highly ordered assembly process leads from the dimers to the IFs. Neither ATP nor GTP is needed for IF assembly, unlike that of microfilaments and microtubules.

IF-associated proteins (IFAPs) mediate the interactions of IFs with one another and with  
5 other cell structures. IFAPs cross-link IFs into a bundle, into a network, or to the plasma membrane, and may cross-link IFs to the microfilament and microtubule cytoskeleton. Microtubules and IFs in particular are closely associated. IFAPs include BPAG1, plakoglobin, desmoplakin I, desmoplakin II, plectin, ankyrin, filaggrin, and lamin B receptor.

The N-terminal portion of ankyrin consists of a repeated 33-amino acid motif, the ankyrin  
10 repeat, which is involved in specific protein-protein interactions. Variable regions within the motif are responsible for specific protein binding, such that different ankyrin repeats are involved in binding to tubulin, anion exchange protein, voltage-gated sodium channel, Na<sup>+</sup>/K<sup>+</sup>-ATPase, and neurofascin. The ankyrin motif is also found in transcription factors, such as NF-κ-B, and in the yeast cell cycle proteins CDC10, SW14, and SW16. Proteins involved in tissue differentiation, such  
15 as Drosophila Notch and C. elegans LIN-12 and GLP-1, also contain ankyrin-like repeats. Lux et al. (1990; Nature 344:36-42) suggest that ankyrin-like repeats function as 'built-in' ankyrins and form binding sites for integral membrane proteins, tubulin, and other proteins.

#### Other Cytoskeleton-Associated Proteins

20 Some cytoskeleton-associated proteins contain a conserved, glycine-rich domain of about 42 residues. This domain, called CAP-Gly, is found in restin, a protein associated with intermediate filaments; vertebrate dynactin, which is associated with dynein; and yeast BIK1 protein which may be required for the formation or stabilization of microtubules during mitosis and for spindle pole body fusion during conjugation. (PROSITE PDOC00660 CAP-Gly domain signature.)

25

#### Proteins of the Erythrocyte Membrane Skeleton

Distribution of oxygen throughout the vertebrate body is effected by red blood cells (erythrocytes). Oxygen diffuses from surrounding water or from the atmosphere through either gill epithelium or pulmonary epithelial type I cells. Oxygen then diffuses through the blood capillary  
30 endothelium directly to the blood circulatory system and through the erythrocyte membrane and is stored as soluble oxyhemoglobin in the cytoplasm. Oxygen is released from hemoglobin at sites throughout the organism and diffuses out from the erythrocyte to other target cells. The structure of the erythrocyte membrane as well as that of other non-erythrocyte cells must be maintained to enable efficient diffusion of oxygen to intracellular compartments.



The erythrocyte membrane is comprised of i) a cholesterol-rich phospholipid bilayer in which many trans-bilayer proteins are embedded, ii) external glycosylphosphatidylinositol-anchored proteins (GPI-proteins), and iii) the erythrocyte or membrane skeleton that laminates the inner surface of the bilayer. The trans-bilayer proteins include anion exchangers, glycophorins, glucose  
 5 transporters, and a variety of cation transporters and pumps. The erythrocyte GPI-proteins include acetylcholinesterase and decay-accelerating factor (CD 55). The skeletal proteins are organized on the cortical, or cytoplasmic, face of the plasma membrane. These proteins include protein 4.1, protein p55,  $\alpha$ - and  $\beta$ -spectrin, actin, and actin-binding proteins such as dematin, tropomyosin, and tropomodulin.  $\alpha$ - and  $\beta$ -spectrin combine to form a heterotetramer in vivo. The spectrin  
 10 heterotetramer organizes into a cortical bidimensional network with a hexagonal mesh. The network is linked to trans-bilayer proteins through a protein complex comprising  $\beta$ -spectrin, ankyrin, anion exchanger, and protein 4.2 and through the "triangular" interaction between protein 4.1, glycophorin C, and protein p55. Structural and functional variants of erythrocyte membrane proteins have been found in a variety of tissues. Variants may be transcribed from multigene families, e.g.,  
 15 anion exchanger, ankyrin, or spectrin, or from single gene families, e.g., protein 4.1 or protein 4.2. mRNA transcripts undergo tissue-specific alternative splicing. Many congenital hemolytic anemias result from mutations in the above-mentioned genes encoding erythrocyte membrane proteins. For example, hereditary elliptocytosis stems from an array of mutations in the spectrin genes at or near the head-to-head self-association region of the spectrin tetramer, or from mutations in the protein 4.1  
 20 gene which reduce levels of protein 4.1. In another example, hereditary spherocytosis is associated with mutations in the ankyrin gene, the anion exchanger gene, the protein 4.2 gene, or the  $\alpha$ - and  $\beta$ -spectrin genes. (Delaunay J. (1995) Transfus. Clin. Biol. 2:207-216.)

Protein 4.1 is an 80 kDa erythrocyte membrane protein with four functional domains. These domains include: i) a 30 kDa basic N-terminal domain, homologous to the ERM  
 25 (Ezrin/Radixin/Moesin) family of actin- and transmembrane protein-binding proteins (Tsukita, S. et al. (1997) Trends Biochem. Sci. 22:53-58); ii) a 16 kDa hydrophilic domain containing a protein kinase C phosphorylation site; iii) a 10 kDa highly charged domain containing a cAMP-dependent protein kinase phosphorylation site critical for the interaction with spectrin and actin; and iv) a 22/24 kDa acidic domain. Protein 4.1 is a member of a structurally and functionally related protein 4.1  
 30 family. The protein 4.1 family is part of an evolutionarily related protein superfamily that includes many tyrosine phosphatases. (Baklouti, F. et al. (1997) Genomics 39:289-302.)

In contrast to the strictly cortical localization of protein 4.1 in mature enucleate erythrocytes, protein 4.1 epitopes have been observed throughout the cytoplasmic compartment and the nucleoskeleton in nucleated cells. In particular, protein 4.1 is present in the nucleoskeleton during

interphase, in the mitotic spindle during mitosis, in perichromatin during telophase, and in the midbody during cytokinesis. (Krauss, S.W. et al. (1997) J. Cell Biol. 137:275-289.)

Differential expression of the protein 4.1 gene resulting in a number of mRNA splice variants has been observed in various human and rodent tissues. Comparison of the gene structure and mRNA  
5 splice variants revealed the extreme genomic sequence conservation of protein 4.1 between different species. The 5' UTR of both the human and rodent mRNA species has not been successfully identified and sequenced, possibly due to GC-rich regions therein which give rise to technical complications during nucleotide sequencing reactions. (Baklouti, supra; Conboy, J.G. (1988) Proc. Natl. Acad. Sci. 85:9062-9065.)

10 Analysis of proteins included in the ERM family of proteins has indicated that the N-terminal domain interacts with intracellular domains of transmembrane proteins such as CD44 and the C-terminal domain binds actin. Both interactions involve interactions with Rho-GTP protein complex, polyphosphoinositides, and serine/threonine kinase and tyrosine kinase activities. Many of the phosphorylation sites on ERM proteins are conserved. Although expression of ERM proteins in vivo  
15 is restricted to tissues such as endothelium, repression of ERM protein gene expression is released under conditions of cell culture. (Tsukita, supra.)

The cortical actin cytoskeleton participates in various membrane-based processes which necessitate a large amount of functional plasticity in the molecular components involved. A family of proteins homologous to band 4.1 is involved in the reorganization of the actin cytoskeleton in  
20 response to various stimuli and probably plays a role in transmembrane signaling. This family includes tyrosine phosphatases, substrates of tyrosine kinases and a candidate for a tumor-suppressor gene. (Arpin M, et al. (1994) Curr. Opin. Cell Biol. 6:136-141.)

Disruptions in cytoskeletal protein interaction have been identified in a number of disease conditions or disorders. Neurofibromatosis type 2 is an autosomal dominant disease of the nervous  
25 system. Schwann cells isolated from patients with neurofibromatosis type 2 have characteristic morphology and growth parameters which differ from control Schwann cells. A gene associated with neurofibromatosis type 2 has been identified and is termed NF2. The NF2 gene product, known as schwannomin or merlin, is a member of the protein 4.1 superfamily, and mutations in the NF2 gene have been shown to be associated with the disease. (Rosenbaum, C. et al. (1998) Neurobiol. Dis.  
30 5:55-64.) In addition, a form of psoriasis may be due to altered expression or distribution in epidermal epithelium of analogs of erythrocyte protein 4.1. (Shimizu, T. (1996) Histol. Histopathol. 11:495-501.) Erythrocytes carrying mutations in spectrin and protein 4.1 showed differing sensitivities to invasion by Plasmodium falciparum. (Facer, C.A. (1995) Parasitol Res. 81:52-57.) Furthermore, antibodies raised against erythrocyte protein 4.1 stained the majority of neurofibrillary

tangles in the prefrontal cortex and hippocampus of brain tissue from patients with Alzheimer's disease. A 68 kDa protein was identified as the most likely brain analog of erythrocyte protein 4.1. (Sihag, R.K. et al. (1994) Brain Res. 656:14-26.)

The discovery of new human cytoskeleton associated proteins and the polynucleotides  
5 encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative, autoimmune/inflammatory, vesicle trafficking, neurological, cell motility, reproductive, and muscle disorders.

### SUMMARY OF THE INVENTION

10 The invention features substantially purified polypeptides, human cytoskeleton associated proteins, referred to collectively as "CYSKP" and individually as "CYSKP-1," "CYSKP-2," "CYSKP-3," "CYSKP-4," "CYSKP-5," "CYSKP-6," "CYSKP-7," "CYSKP-8," "CYSKP-9," "CYSKP-10," "CYSKP-11," "CYSKP-12," "CYSKP-13," "CYSKP-14," "CYSKP-15," and "CYSKP-16." In one aspect, the invention provides a substantially purified polypeptide comprising  
15 an amino acid sequence selected from the group consisting of SEQ ID NO:1-16 and fragments thereof.

The invention further provides a substantially purified variant having at least 90% amino acid identity to at least one of the amino acid sequences selected from the group consisting of SEQ ID NO:1-16 and fragments thereof. The invention also provides an isolated and purified polynucleotide  
20 encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-16 and fragments thereof. The invention also includes an isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-16 and fragments thereof.

25 Additionally, the invention provides an isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-16 and fragments thereof. The invention also provides an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide encoding the polypeptide comprising the amino acid sequence  
30 selected from the group consisting of SEQ ID NO:1-16 and fragments thereof.

The invention also provides a method for detecting a polynucleotide in a sample containing nucleic acids, the method comprising the steps of (a) hybridizing the complement of the polynucleotide sequence to at least one of the polynucleotides of the sample, thereby forming a hybridization complex; and (b) detecting the hybridization complex, wherein the presence of the

hybridization complex correlates with the presence of a polynucleotide in the sample. In one aspect, the method further comprises amplifying the polynucleotide prior to hybridization.

The invention also provides an isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:17-32, and fragments thereof. The invention further provides an isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide sequence selected from the group consisting of SEQ ID NO:17-32 and fragments thereof. The invention also provides an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:17-32 and fragments thereof.

The invention further provides an expression vector containing at least a fragment of the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-16 and fragments thereof. In another aspect, the expression vector is contained within a host cell.

The invention also provides a method for producing a polypeptide, the method comprising the steps of: (a) culturing the host cell containing an expression vector containing at least a fragment of a polynucleotide under conditions suitable for the expression of the polypeptide; and (b) recovering the polypeptide from the host cell culture.

The invention also provides a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:1-16 and fragments thereof, in conjunction with a suitable pharmaceutical carrier.

The invention further includes a purified antibody which binds to a polypeptide selected from the group consisting of SEQ ID NO:1-16 and fragments thereof. The invention also provides a purified agonist and a purified antagonist to the polypeptide.

The invention also provides a method for treating or preventing a disorder associated with decreased expression or activity of CYSKP, the method comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:1-16 and fragments thereof, in conjunction with a suitable pharmaceutical carrier.

The invention also provides a method for treating or preventing a disorder associated with increased expression or activity of CYSKP, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-16 and fragments thereof.

### BRIEF DESCRIPTION OF THE TABLES

Table 1 shows polypeptide and nucleotide sequence identification numbers (SEQ ID NOs), clone identification numbers (clone IDs), cDNA libraries, and cDNA fragments used to assemble full-length sequences encoding CYSKP.

Table 2 shows features of each polypeptide sequence, including potential motifs, homologous sequences, and methods and algorithms used for identification of CYSKP.

Table 3 shows selected fragments of each nucleic acid sequence; the tissue-specific expression patterns of each nucleic acid sequence as determined by northern analysis; diseases, disorders, or conditions associated with these tissues; and the vector into which each cDNA was cloned.

Table 4 describes the tissues used to construct the cDNA libraries from which cDNA clones encoding CYSKP were isolated.

Table 5 shows the tools, programs, and algorithms used to analyze CYSKP, along with applicable descriptions, references, and threshold parameters.

### DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

## DEFINITIONS

“CYSKP” refers to the amino acid sequences of substantially purified CYSKP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and preferably the human species, from any source, whether natural, synthetic, semi-synthetic, or  
5 recombinant.

The term “agonist” refers to a molecule which, when bound to CYSKP, increases or prolongs the duration of the effect of CYSKP. Agonists may include proteins, nucleic acids, carbohydrates, or any other molecules which bind to and modulate the effect of CYSKP.

An “allelic variant” is an alternative form of the gene encoding CYSKP. Allelic variants may  
10 result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. Any given natural or recombinant gene may have none, one, or many allelic forms. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one  
15 or more times in a given sequence.

“Altered” nucleic acid sequences encoding CYSKP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polynucleotide the same as CYSKP or a polypeptide with at least one functional characteristic of CYSKP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide  
20 probe of the polynucleotide encoding CYSKP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding CYSKP. The encoded protein may also be “altered,” and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent CYSKP. Deliberate amino acid substitutions may be made on the basis of similarity in  
25 polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of CYSKP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, positively charged amino acids may include lysine and arginine, and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine; glycine and alanine;  
30 asparagine and glutamine; serine and threonine; and phenylalanine and tyrosine.

The terms “amino acid” and “amino acid sequence” refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. In this context, “fragments,” “immunogenic fragments,” or “antigenic fragments” refer to fragments of CYSKP which are preferably at least 5 to about 15 amino acids in

length, most preferably at least 14 amino acids, and which retain some biological activity or immunological activity of CYSKP. Where "amino acid sequence" is recited to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with  
5 the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which, when bound to CYSKP, decreases the  
10 amount or the duration of the effect of the biological or immunological activity of CYSKP. Antagonists may include proteins, nucleic acids, carbohydrates, antibodies, or any other molecules which decrease the effect of CYSKP.

The term "antibody" refers to intact molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding the epitopic determinant. Antibodies that  
15 bind CYSKP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and  
20 keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that fragment of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (given regions or three-dimensional structures on  
25 the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition containing a nucleic acid sequence which is complementary to the "sense" strand of a specific nucleic acid sequence. Antisense molecules may be produced by any method including synthesis or transcription. Once introduced into a cell, the  
30 complementary nucleotides combine with natural sequences produced by the cell to form duplexes and to block either transcription or translation. The designation "negative" can refer to the antisense strand, and the designation "positive" can refer to the sense strand.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the

capability of the natural, recombinant, or synthetic CYSKP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" and "complementarity" refer to the natural binding of polynucleotides by base pairing. For example, the sequence "5' A-G-T 3'" bonds to the  
5 complementary sequence "3' T-C-A 5'." Complementarity between two single-stranded molecules may be "partial," such that only some of the nucleic acids bind, or it may be "complete," such that total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands. This is of particular importance in amplification reactions, which  
10 depend upon binding between nucleic acids strands, and in the design and use of peptide nucleic acid (PNA) molecules.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution.  
15 Compositions comprising polynucleotide sequences encoding CYSKP or fragments of CYSKP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).  
20 "Consensus sequence" refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, extended using the XL-PCR kit (Perkin-Elmer, Norwalk CT) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from the overlapping sequences of more than one Incyte Clone using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI). Some sequences have been both extended and  
25 assembled to produce the consensus sequence.

The term "correlates with expression of a polynucleotide" indicates that the detection of the presence of nucleic acids, the same or related to a nucleic acid sequence encoding CYSKP, by northern analysis is indicative of the presence of nucleic acids encoding CYSKP in a sample, and thereby correlates with expression of the transcript from the polynucleotide encoding CYSKP.

30 A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to the chemical modification of a polypeptide sequence, or a polynucleotide sequence. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, or amino group. A derivative polynucleotide



encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

5       The term "similarity" refers to a degree of complementarity. There may be partial similarity or complete similarity. The word "identity" may substitute for the word "similarity." A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to as "substantially similar." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization  
10 assay (Southern or northern blot, solution hybridization, and the like) under conditions of reduced stringency. A substantially similar sequence or hybridization probe will compete for and inhibit the binding of a completely similar (identical) sequence to the target sequence under conditions of reduced stringency. This is not to say that conditions of reduced stringency are such that non-specific binding is permitted, as reduced stringency conditions require that the binding of two sequences to  
15 one another be a specific (i.e., a selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% similarity or identity). In the absence of non-specific binding, the substantially similar sequence or probe will not hybridize to the second non-complementary target sequence.

20       The phrases "percent identity" and "% identity" refer to the percentage of sequence similarity found in a comparison of two or more amino acid or nucleic acid sequences. Percent identity can be determined electronically, e.g., by using the MEGALIGN program (DNASTAR, Madison WI) which creates alignments between two or more sequences according to methods selected by the user, e.g., the clustal method. (See, e.g., Higgins, D.G. and P.M. Sharp (1988) *Gene* 73:237-244.) Parameters  
25 for each method may be the default parameters provided by MEGALIGN or may be specified by the user. The clustal algorithm groups sequences into clusters by examining the distances between all pairs. The clusters are aligned pairwise and then in groups. The percentage similarity between two amino acid sequences, e.g., sequence A and sequence B, is calculated by dividing the length of sequence A, minus the number of gap residues in sequence A, minus the number of gap residues in  
30 sequence B, into the sum of the residue matches between sequence A and sequence B, times one hundred. Gaps of low or of no similarity between the two amino acid sequences are not included in determining percentage similarity. Percent identity between nucleic acid sequences can also be counted or calculated by other methods known in the art, e.g., the Jotun Hein method. (See, e.g., Hein, J. (1990) *Methods Enzymol.* 183:626-645.) Identity between sequences can also be determined

by other methods known in the art, e.g., by varying hybridization conditions.

“Human artificial chromosomes” (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for stable mitotic chromosome segregation and maintenance.

5       The term “humanized antibody” refers to antibody molecules in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

“Hybridization” refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

10       The term “hybridization complex” refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g.,  $C_0t$  or  $R_0t$  analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate  
15 to which cells or their nucleic acids have been fixed).

The words “insertion” and “addition” refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively, to the sequence found in the naturally occurring molecule.

“Immune response” can refer to conditions associated with inflammation, trauma, immune  
20 disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

The term “microarray” refers to an arrangement of distinct polynucleotides on a substrate.

The terms “element” and “array element” in a microarray context, refer to hybridizable  
25 polynucleotides arranged on the surface of a substrate.

The term “modulate” refers to a change in the activity of CYSKP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of CYSKP.

The phrases “nucleic acid” or “nucleic acid sequence,” as used herein, refer to a nucleotide,  
30 oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material. In this context, “fragments” refers to those nucleic acid sequences which comprise a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:17-32, for example, as

distinct from any other sequence in the same genome. For example, a fragment of SEQ ID NO:17-32 is useful in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:17-32 from related polynucleotide sequences. A fragment of SEQ ID NO:17-32 is at least about 15-20 nucleotides in length. The precise length of the fragment of SEQ ID NO:17-32 and the region of SEQ ID NO:17-32 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment. In some cases, a fragment, when translated, would produce polypeptides retaining some functional characteristic, e.g., antigenicity, or structural domain characteristic, e.g., ATP-binding site, of the full-length polypeptide.

The terms "operably associated" and "operably linked" refer to functionally related nucleic acid sequences. A promoter is operably associated or operably linked with a coding sequence if the promoter controls the translation of the encoded polypeptide. While operably associated or operably linked nucleic acid sequences can be contiguous and in the same reading frame, certain genetic elements, e.g., repressor genes, are not contiguously linked to the sequence encoding the polypeptide but still bind to operator sequences that control expression of the polypeptide.

The term "oligonucleotide" refers to a nucleic acid sequence of at least about 6 nucleotides to 60 nucleotides, preferably about 15 to 30 nucleotides, and most preferably about 20 to 25 nucleotides, which can be used in PCR amplification or in a hybridization assay or microarray. "Oligonucleotide" is substantially equivalent to the terms "amplimer," "primer," "oligomer," and "probe," as these terms are commonly defined in the art.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

The term "sample" is used in its broadest sense. A sample suspected of containing nucleic acids encoding CYSKP, or fragments thereof, or CYSKP itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, or an antagonist. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide containing the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the

antibody.

The term "stringent conditions" refers to conditions which permit hybridization between polynucleotides and the claimed polynucleotides. Stringent conditions can be defined by salt concentration, the concentration of organic solvent, e.g., formamide, temperature, and other  
5 conditions well known in the art. In particular, stringency can be increased by reducing the concentration of salt, increasing the concentration of formamide, or raising the hybridization temperature.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free,  
10 preferably about 75% free, and most preferably about 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters,  
15 chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

"Transformation" describes a process by which exogenous DNA enters and changes a recipient cell. Transformation may occur under natural or artificial conditions according to various  
20 methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of  
25 replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "variant" of CYSKP polypeptides refers to an amino acid sequence that is altered by one or more amino acid residues. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties (e.g., replacement of leucine with isoleucine).  
30 More rarely, a variant may have "nonconservative" changes (e.g., replacement of glycine with tryptophan). Analogous minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, LASERGENE software (DNASTAR).

The term "variant," when used in the context of a polynucleotide sequence, may encompass a polynucleotide sequence related to CYSKP. This definition may also include, for example, "allelic" (as defined above), "splice," "species," or "polymorphic" variants. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or an absence of domains. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

#### THE INVENTION

The invention is based on the discovery of new human cytoskeleton associated proteins (CYSKP), the polynucleotides encoding CYSKP, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, autoimmune/inflammatory, vesicle trafficking, neurological, cell motility, reproductive, and muscle disorders.

Table 1 lists the Incyte clones used to assemble full length nucleotide sequences encoding CYSKP. Columns 1 and 2 show the sequence identification numbers (SEQ ID NOs) of the polypeptide and nucleotide sequences, respectively. Column 3 shows the clone IDs of the Incyte clones in which nucleic acids encoding each CYSKP were identified, and column 4 shows the cDNA libraries from which these clones were isolated. Column 5 shows Incyte clones and their corresponding cDNA libraries. Clones for which cDNA libraries are not indicated were derived from pooled cDNA libraries. The clones in column 5 were used to assemble the consensus nucleotide sequence of each CYSKP and are useful as fragments in hybridization technologies.

The columns of Table 2 show various properties of each of the polypeptides of the invention: column 1 references the SEQ ID NO; column 2 shows the number of amino acid residues in each polypeptide; column 3 shows potential phosphorylation sites; column 4 shows potential glycosylation sites; column 5 shows the amino acid residues comprising signature sequences and motifs; column 6 shows homologous sequences as identified by BLAST analysis; and column 7 shows analytical methods used to characterize each polypeptide through sequence homology and protein motifs.

The columns of Table 3 show the tissue-specificity and diseases, disorders, or conditions associated with nucleotide sequences encoding CYSKP. The first column of Table 3 lists the nucleotide SEQ ID NOs. Column 2 lists fragments of the nucleotide sequences of column 1. These

fragments are useful, for example, in hybridization or amplification technologies to identify SEQ ID NO:17-32 and to distinguish between SEQ ID NO:17-32 and related polynucleotide sequences. The polypeptides encoded by these fragments are useful, for example, as immunogenic peptides. Column 3 lists tissue categories which express CYSKP as a fraction of total tissues expressing CYSKP.

- 5 Column 4 lists diseases, disorders, or conditions associated with those tissues expressing CYSKP as a fraction of total tissues expressing CYSKP. Column 5 lists the vectors used to subclone each cDNA library.

Of particular note is the expression of SEQ ID NO:31 in nervous tissues and the expression of SEQ ID NO:32 in musculoskeletal tissues.

- 10 The columns of Table 4 show descriptions of the tissues used to construct the cDNA libraries from which cDNA clones encoding CYSKP were isolated. Column 1 references the nucleotide SEQ ID NOs, column 2 shows the cDNA libraries from which these clones were isolated, and column 3 shows the tissue origins and other descriptive information relevant to the cDNA libraries in column 2.

- The invention also encompasses CYSKP variants. A preferred CYSKP variant is one which  
15 has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the CYSKP amino acid sequence, and which contains at least one functional or structural characteristic of CYSKP.

- The invention also encompasses polynucleotides which encode CYSKP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected  
20 from the group consisting of SEQ ID NO:17-32, which encodes CYSKP.

- The invention also encompasses a variant of a polynucleotide sequence encoding CYSKP. In particular, such a variant polynucleotide sequence will have at least about 70%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding CYSKP. A particular aspect of the invention encompasses a  
25 variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:17-32 which has at least about 70%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:17-32. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural  
30 characteristic of CYSKP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding CYSKP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide

sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring CYSKP, and all such variations are to be considered as being specifically disclosed.

5           Although nucleotide sequences which encode CYSKP and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring CYSKP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding CYSKP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the  
10 peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding CYSKP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

15           The invention also encompasses production of DNA sequences which encode CYSKP and CYSKP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding CYSKP or any fragment thereof.

20           Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:17-32 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) For example, stringent salt concentration will ordinarily be less than about 750 mM  
25 NaCl and 75 mM trisodium citrate, preferably less than about 500 mM NaCl and 50 mM trisodium citrate, and most preferably less than about 250 mM NaCl and 25 mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent, e.g., formamide, while high stringency hybridization can be obtained in the presence of at least about 35% formamide, and most preferably at least about 50% formamide. Stringent temperature conditions will ordinarily  
30 include temperatures of at least about 30°C, more preferably of at least about 37°C, and most preferably of at least about 42°C. Varying additional parameters, such as hybridization time, the concentration of detergent, e.g., sodium dodecyl sulfate (SDS), and the inclusion or exclusion of carrier DNA, are well known to those skilled in the art. Various levels of stringency are accomplished by combining these various conditions as needed. In a preferred embodiment,

hybridization will occur at 30°C in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37°C in 500 mM NaCl, 50 mM trisodium citrate, 1% SDS, 35% formamide, and 100 µg/ml denatured salmon sperm DNA (ssDNA). In a most preferred embodiment, hybridization will occur at 42°C in 250 mM NaCl, 25 mM trisodium citrate, 1% SDS, 50 % formamide, and 200 µg/ml ssDNA. Useful variations on these conditions will be readily apparent to those skilled in the art.

The washing steps which follow hybridization can also vary in stringency. Wash stringency conditions can be defined by salt concentration and by temperature. As above, wash stringency can be increased by decreasing salt concentration or by increasing temperature. For example, stringent salt concentration for the wash steps will preferably be less than about 30 mM NaCl and 3 mM trisodium citrate, and most preferably less than about 15 mM NaCl and 1.5 mM trisodium citrate. Stringent temperature conditions for the wash steps will ordinarily include temperature of at least about 25°C, more preferably of at least about 42°C, and most preferably of at least about 68°C. In a preferred embodiment, wash steps will occur at 25°C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 42°C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. In a most preferred embodiment, wash steps will occur at 68°C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. Additional variations on these conditions will be readily apparent to those skilled in the art.

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Perkin-Elmer), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Perkin-Elmer). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Perkin-Elmer), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding CYSKP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences,



such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-306.) Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Perkin-Elmer), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof

which encode CYSKP may be cloned in recombinant DNA molecules that direct expression of CYSKP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express CYSKP.

5       The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter CYSKP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-  
10 mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

In another embodiment, sequences encoding CYSKP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucl. Acids Res. Symp. Ser. 7:215-223, and Horn, T. et al. (1980) Nucl. Acids Res. Symp. Ser. 7:225-232.)  
15 Alternatively, CYSKP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solid-phase techniques. (See, e.g., Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Perkin-Elmer). Additionally, the amino acid sequence of CYSKP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other  
20 proteins, or any part thereof, to produce a variant polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH  
25 Freeman, New York NY.)

In order to express a biologically active CYSKP, the nucleotide sequences encoding CYSKP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers,  
30 constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding CYSKP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding CYSKP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding CYSKP and its initiation codon and

upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and  
5 initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding CYSKP and appropriate transcriptional and translational  
10 control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

15 A variety of expression vector/host systems may be utilized to contain and express sequences encoding CYSKP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower  
20 mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding CYSKP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding CYSKP can be achieved using a  
25 multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or pSPORT1 plasmid (Life Technologies). Ligation of sequences encoding CYSKP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for  
30 in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509.) When large quantities of CYSKP are needed, e.g. for the production of antibodies, vectors which direct high level expression of CYSKP may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of CYSKP. A number of vectors

containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 5 1995, supra; Grant et al. (1987) *Methods Enzymol.* 153:516-54; and Scorer, C. A. et al. (1994) *Bio/Technology* 12:181-184.)

Plant systems may also be used for expression of CYSKP. Transcription of sequences encoding CYSKP may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J.* 10 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; and Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology 15 (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding CYSKP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain 20 infective virus which expresses CYSKP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci.* 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of 25 DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression 30 of CYSKP in cell lines is preferred. For example, sequences encoding CYSKP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to

confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These  
5 include, but are not limited to, the herpes simplex virus thymidine kinase and adenine  
phosphoribosyltransferase genes, for use in *tk* or *ap<sup>r</sup>* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* or *pat*  
10 confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins  
15 (GFP; Clontech),  $\beta$  glucuronidase and its substrate  $\beta$ -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is  
20 also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding CYSKP is inserted within a marker gene sequence, transformed cells containing sequences encoding CYSKP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding CYSKP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually  
25 indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding CYSKP and that express CYSKP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR  
amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or  
30 chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of CYSKP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing

monoclonal antibodies reactive to two non-interfering epitopes on CYSKP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St Paul MN, Sect. IV; Coligan, J. E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding CYSKP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding CYSKP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding CYSKP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode CYSKP may be designed to contain signal sequences which direct secretion of CYSKP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid

sequences encoding CYSKP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric CYSKP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of CYSKP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the CYSKP encoding sequence and the heterologous protein sequence, so that CYSKP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled CYSKP may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract systems (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, preferably <sup>35</sup>S-methionine.

Fragments of CYSKP may be produced not only by recombinant production, but also by direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, supra, pp. 55-60.) Protein synthesis may be performed by manual techniques or by automation. Automated synthesis may be achieved, for example, using the ABI 431A peptide synthesizer (Perkin-Elmer). Various fragments of CYSKP may be synthesized separately and then combined to produce the full length molecule.

## THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of CYSKP and human cytoskeleton associated proteins. In addition, the expression of CYSKP is closely associated with cancer, cell proliferation, inflammation, immune response, musculoskeletal, nervous, reproductive, cardiovascular, and gastrointestinal tissues. Therefore, CYSKP appears to play a role in cell proliferative, autoimmune/inflammatory, vesicle trafficking,

neurological, cell motility, reproductive, and muscle disorders. In the treatment of disorders associated with increased CYSKP expression or activity, it is desirable to decrease the expression or activity of CYSKP. In the treatment of disorders associated with decreased CYSKP expression or activity, it is desirable to increase the expression or activity of CYSKP.

- 5 Therefore, in one embodiment, CYSKP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CYSKP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria,
- 10 polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an autoimmune/inflammatory
- 15 disorder such as acquired immunodeficiency syndrome (AIDS), actinic keratosis, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, arteriosclerosis, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, bursitis, cholecystitis, cirrhosis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis,
- 20 diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, paroxysmal nocturnal hemoglobinuria, hepatitis, hypereosinophilia, irritable bowel syndrome, mixed connective tissue disease (MCTD), myelofibrosis, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis,
- 25 polycythemia vera, polymyositis, primary thrombocythemia, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a vesicle trafficking disorder such
- 30 as cystic fibrosis, glucose-galactose malabsorption syndrome, hypercholesterolemia, diabetes mellitus, diabetes insipidus, hyper- and hypoglycemia, Grave's disease, goiter, Cushing's disease, and Addison's disease, gastrointestinal disorders including ulcerative colitis, gastric and duodenal ulcers, other conditions associated with abnormal vesicle trafficking, including acquired immunodeficiency syndrome (AIDS), allergies including hay fever, asthma, and urticaria (hives), autoimmune hemolytic



anemia, proliferative glomerulonephritis, inflammatory bowel disease, multiple sclerosis, myasthenia gravis, rheumatoid and osteoarthritis, scleroderma, Chediak-Higashi and Sjogren's syndromes, systemic lupus erythematosus, toxic shock syndrome, traumatic tissue damage, and viral, bacterial, fungal, helminthic, and protozoal infections; a neurological disorder such as epilepsy, ischemic

5 cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial

10 thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy,

15 neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis; inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy,

20 tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a cell motility disorder such as ankylosing spondylitis, Chediak-Higashi syndrome, Duchenne and Becker muscular dystrophy, intrahepatic cholestasis, myocardial hyperplasia, cardiomyopathy, early onset periodontitis, cancers such as adenocarcinoma, ovarian carcinoma, and chronic myelogenous

25 leukemia, and bacterial and helminthic infections; and a heart and skeletal muscle disorder such as cardiomyopathy, myocarditis, Duchenne's muscular dystrophy, Becker's muscular dystrophy, myotonic dystrophy, central core disease, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, infectious myositis, polymyositis, dermatomyositis, inclusion body myositis, thyrotoxic myopathy, and ethanol myopathy.

30 In another embodiment, a vector capable of expressing CYSKP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CYSKP including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified CYSKP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat

or prevent a disorder associated with decreased expression or activity of CYSKP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of CYSKP may be administered to a subject to treat or prevent a disorder associated with decreased expression or  
5 activity of CYSKP including, but not limited to, those listed above.

In a further embodiment, an antagonist of CYSKP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CYSKP. Examples of such disorders include, but are not limited to, those cell proliferative, autoimmune/inflammatory, vesicle trafficking, neurological, cell motility, and heart and skeletal muscle disorders described above; a  
10 reproductive disorder such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, and endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian tumors, uterine fibroids, autoimmune disorders, ectopic pregnancies, and teratogenesis, cancer of the breast, fibrocystic breast disease, and galactorrhea, a disruption of spermatogenesis, abnormal sperm  
15 physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia; and a smooth muscle disorder. A smooth muscle disorder is defined as any impairment or alteration in the normal action of smooth muscle and may include, but is not limited to, angina, anaphylactic shock, arrhythmias, asthma, cardiovascular shock, Cushing's syndrome, hypertension, hypoglycemia, myocardial  
20 infarction, migraine, and pheochromocytoma, and myopathies including cardiomyopathy, encephalopathy, epilepsy, Kearns-Sayre syndrome, lactic acidosis, myoclonic disorder, and ophthalmoplegia. Smooth muscle includes, but is not limited to, that of the blood vessels, gastrointestinal tract, heart, and uterus. In one aspect, an antibody which specifically binds CYSKP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a  
25 pharmaceutical agent to cells or tissue which express CYSKP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding CYSKP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CYSKP including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary  
30 sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic

efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of CYSKP may be produced using methods which are generally known in the art. In particular, purified CYSKP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind CYSKP. Antibodies to CYSKP may  
5 also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans,  
10 and others may be immunized by injection with CYSKP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in  
15 humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to CYSKP have an amino acid sequence consisting of at least about 5 amino acids, and, more preferably, of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein and contain the  
20 entire amino acid sequence of a small, naturally occurring molecule. Short stretches of CYSKP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to CYSKP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not  
25 limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the  
30 splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce CYSKP-specific single

chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton D.R. (1991) Proc. Natl. Acad. Sci. 88:10134-10137.)

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte  
5 population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. 86: 3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for CYSKP may also be generated. For example, such fragments include, but are not limited to, F(ab')<sub>2</sub> fragments produced by pepsin  
10 digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired  
15 specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between CYSKP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering CYSKP epitopes is preferred, but a competitive binding assay may  
20 also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for CYSKP. Affinity is expressed as an association constant,  $K_a$ , which is defined as the molar concentration of CYSKP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions.  
25 The  $K_a$  determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple CYSKP epitopes, represents the average affinity, or avidity, of the antibodies for CYSKP. The  $K_a$  determined for a preparation of monoclonal antibodies, which are monospecific for a particular CYSKP epitope, represents a true measure of affinity. High-affinity antibody preparations with  $K_a$  ranging from about  $10^9$  to  $10^{12}$  L/mole are preferred for use in immunoassays in  
30 which the CYSKP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with  $K_a$  ranging from about  $10^6$  to  $10^7$  L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of CYSKP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington, DC; Liddell, J. E. and Cryer, A. (1991) A Practical Guide to

Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is preferred for use in procedures requiring precipitation of CYSKP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding CYSKP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding CYSKP may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding CYSKP. Thus, complementary molecules or fragments may be used to modulate CYSKP activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding CYSKP.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. Methods which are well known to those skilled in the art can be used to construct vectors to express nucleic acid sequences complementary to the polynucleotides encoding CYSKP. (See, e.g., Sambrook, supra; Ausubel, 1995, supra.)

Genes encoding CYSKP can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide, or fragment thereof, encoding CYSKP. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector, and may last even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5', or regulatory regions of the gene encoding CYSKP. Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful

because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-  
5 177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example,  
10 engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding CYSKP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides,  
15 corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared  
20 by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding CYSKP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA  
25 constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase  
30 linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nature Biotechnology 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

10 An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of CYSKP, antibodies to CYSKP, and mimetics, agonists, antagonists, or inhibitors of CYSKP. The compositions may be administered alone or in combination with at least one other agent, such as a  
15 stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial,  
20 intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on  
25 techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees,  
30 capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combining active compounds with solid excipient and processing the resultant mixture of granules (optionally, after grinding) to obtain tablets or dragee cores. Suitable auxiliaries can be added, if desired. Suitable excipients include carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol,

and sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums, including arabic and tragacanth; and proteins, such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, and alginic acid or a salt thereof, 5 such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to 10 characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, 15 the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances 20 which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils, such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate, triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain suitable 25 stabilizers or agents to increase the solubility of the compounds and allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner 30 that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acids. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding



free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1 mM to 50 mM histidine, 0.1% to 2% sucrose, and 2% to 7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of CYSKP, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells or in animal models such as mice, rats, rabbits, dogs, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example CYSKP or fragments thereof, antibodies of CYSKP, and agonists, antagonists or inhibitors of CYSKP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the  $ED_{50}$  (the dose therapeutically effective in 50% of the population) or  $LD_{50}$  (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the  $LD_{50}/ED_{50}$  ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the  $ED_{50}$  with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1  $\mu\text{g}$  to 100,000  $\mu\text{g}$ , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their  
5 inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

## DIAGNOSTICS

In another embodiment, antibodies which specifically bind CYSKP may be used for the diagnosis of disorders characterized by expression of CYSKP, or in assays to monitor patients being  
10 treated with CYSKP or agonists, antagonists, or inhibitors of CYSKP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for CYSKP include methods which utilize the antibody and a label to detect CYSKP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter  
15 molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring CYSKP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of CYSKP expression. Normal or standard values for CYSKP expression are established by combining body fluids or cell  
20 extracts taken from normal mammalian subjects, preferably human, with antibody to CYSKP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, preferably by photometric means. Quantities of CYSKP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

25 In another embodiment of the invention, the polynucleotides encoding CYSKP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which expression of CYSKP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess  
30 expression of CYSKP, and to monitor regulation of CYSKP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding CYSKP or closely related molecules may be used to identify nucleic acid sequences which encode CYSKP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a

conserved motif, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low), will determine whether the probe identifies only naturally occurring sequences encoding CYSKP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably have at least 50% sequence identity to any of the CYSKP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:17-32 or from genomic sequences including promoters, enhancers, and introns of the CYSKP gene.

Means for producing specific hybridization probes for DNAs encoding CYSKP include the cloning of polynucleotide sequences encoding CYSKP or CYSKP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as <sup>32</sup>P or <sup>35</sup>S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding CYSKP may be used for the diagnosis of disorders associated with expression of CYSKP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), actinic keratosis, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, arteriosclerosis, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, bursitis, cholecystitis, cirrhosis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, paroxysmal nocturnal hemoglobinuria, hepatitis, hypereosinophilia, irritable bowel syndrome, mixed connective tissue disease (MCTD), myelofibrosis, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polycythemia vera, polymyositis, primary

thrombocythemia, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and

5 trauma; a vesicle trafficking disorder such as cystic fibrosis, glucose-galactose malabsorption syndrome, hypercholesterolemia, diabetes mellitus, diabetes insipidus, hyper- and hypoglycemia, Grave's disease, goiter, Cushing's disease, and Addison's disease, gastrointestinal disorders including ulcerative colitis, gastric and duodenal ulcers, other conditions associated with abnormal vesicle trafficking, including acquired immunodeficiency syndrome (AIDS), allergies including hay fever,

10 asthma, and urticaria (hives), autoimmune hemolytic anemia, proliferative glomerulonephritis, inflammatory bowel disease, multiple sclerosis, myasthenia gravis, rheumatoid and osteoarthritis, scleroderma, Chediak-Higashi and Sjogren's syndromes, systemic lupus erythematosus, toxic shock syndrome, traumatic tissue damage, and viral, bacterial, fungal, helminthic, and protozoal infections; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral

15 neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central

20 nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial

25 nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic

30 neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a cell motility disorder such as ankylosing spondylitis, Chediak-Higashi syndrome, Duchenne and Becker muscular dystrophy, intrahepatic cholestasis, myocardial hyperplasia, cardiomyopathy, early onset periodontitis, cancers such as adenocarcinoma, ovarian carcinoma, and chronic myelogenous leukemia, and bacterial and helminthic infections; a heart and

skeletal muscle disorder such as cardiomyopathy, myocarditis, Duchenne's muscular dystrophy, Becker's muscular dystrophy, myotonic dystrophy, central core disease, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, infectious myositis, polymyositis, dermatomyositis, inclusion body myositis, thyrotoxic myopathy, and ethanol myopathy; a

5 reproductive disorder such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, and endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian tumors, uterine fibroids, autoimmune disorders, ectopic pregnancies, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea, a disruption of spermatogenesis, abnormal sperm

10 physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia; and a smooth muscle disorder. A smooth muscle disorder is defined as any impairment or alteration in the normal action of smooth muscle and may include, but is not limited to, angina, anaphylactic shock, arrhythmias, asthma, cardiovascular shock, Cushing's syndrome, hypertension, hypoglycemia, myocardial

15 infarction, migraine, and pheochromocytoma, and myopathies including cardiomyopathy, encephalopathy, epilepsy, Kearns-Sayre syndrome, lactic acidosis, myoclonic disorder, and ophthalmoplegia. Smooth muscle includes, but is not limited to, that of the blood vessels, gastrointestinal tract, heart, and uterus. The polynucleotide sequences encoding CYSKP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR

20 technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered CYSKP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding CYSKP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide

25 sequences encoding CYSKP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding CYSKP in the

30 sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of CYSKP, a normal or standard profile for expression is established. This may be accomplished by

combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding CYSKP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified  
5 polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the  
10 patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the  
15 development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding  
20 CYSKP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding CYSKP, or a fragment of a polynucleotide complementary to the polynucleotide encoding CYSKP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or  
25 quantitation of closely related DNA or RNA sequences.

Methods which may also be used to quantify the expression of CYSKP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be  
30 accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to identify

genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.)

In another embodiment of the invention, nucleic acid sequences encoding CYSKP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) site. Correlation between the location of the gene encoding CYSKP on a physical chromosomal map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder. The nucleotide sequences of the invention may be used to detect differences in gene sequences among normal, carrier, and affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion,

etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, CYSKP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between CYSKP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with CYSKP, or fragments thereof, and washed. Bound CYSKP is then detected by methods well known in the art. Purified CYSKP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding CYSKP specifically compete with a test compound for binding CYSKP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with CYSKP.

In additional embodiments, the nucleotide sequences which encode CYSKP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. 60/131,321, and [Atty Docket No. PF-0594 P, filed September 18, 1998] are hereby expressly incorporated by reference.



## EXAMPLES

### I. Construction of cDNA Libraries

RNA was purchased from Clontech or isolated from tissues described in Table 4. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A+) RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIP<sup>T</sup> plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), pSPORT1 plasmid (Life Technologies), or pINCY (Incyte Pharmaceuticals, Palo Alto CA). Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or ElectroMAX DH10B from Life Technologies.

### II. Isolation of cDNA Clones

Plasmids were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8

Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

### 10 III. Sequencing and Analysis

cDNA sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Perkin-Elmer) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Perkin-Elmer) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example V.

The polynucleotide sequences derived from cDNA sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those skilled in the art. Table 5 summarizes the tools, programs, and algorithms used and provides applicable descriptions, references, and threshold parameters. The first column of Table 5 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score, the greater the homology between two sequences). Sequences were analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments were generated using the default

parameters specified by the clustal algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

The polynucleotide sequences were validated by removing vector, linker, and polyA sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS to acquire annotation using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length amino acid sequences, and these full length sequences were subsequently analyzed by querying against databases such as the GenBank databases (described above), SwissProt, BLOCKS, PRINTS, Prosite, and Hidden Markov Model (HMM)-based protein family databases such as PFAM. HMM is a probabilistic approach which analyzes consensus primary structures of gene families. (See, e.g., Eddy, S.R. (1996) Curr. Opin. Str. Biol. 6:361-365.)

The programs described above for the assembly and analysis of full length polynucleotide and amino acid sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:17-32. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies were described in The Invention section above.

#### IV. Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, *supra*, ch. 7; Ausubel, 1995, *supra*, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in nucleotide databases such as GenBank or LIFESEQ (Incyte Pharmaceuticals). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\% \text{ sequence identity} \times \% \text{ maximum BLAST score}}{100}$$

100

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1% to 2% error, and, with a product score of 70, the match will be exact. Similar molecules

are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

The results of northern analyses are reported as a percentage distribution of libraries in which the transcript encoding CYSKP occurred. Analysis involved the categorization of cDNA libraries by organ/tissue and disease. The organ/tissue categories included cardiovascular, dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous, reproductive, and urologic. The disease/condition categories included cancer, inflammation/trauma, cell proliferation, neurological, and pooled. For each category, the number of libraries expressing the sequence of interest was counted and divided by the total number of libraries across all categories. Percentage values of tissue-specific and disease- or condition-specific expression are reported in Table 3.

#### V. Extension of CYSKP Encoding Polynucleotides

The full length nucleic acid sequences of SEQ ID NO:17-32 were produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer, to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing  $Mg^{2+}$ ,  $(NH_4)_2SO_4$ , and  $\beta$ -mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100  $\mu$ l PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5  $\mu$ l of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar,

Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5  $\mu$ l to 10  $\mu$ l aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose mini-gel to determine which reactions were successful in extending  
5 the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%)  
10 agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, individual colonies were picked and cultured overnight at 37°C in 384-  
15 well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was  
20 quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulphoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer).

25 In like manner, the nucleotide sequences of SEQ ID NO:17-32 are used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for such extension, and an appropriate genomic library.

#### **VI. Labeling and Use of Individual Hybridization Probes**

Hybridization probes derived from SEQ ID NO:17-32 are employed to screen cDNAs,  
30 genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of [ $\gamma$ -<sup>32</sup>P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase  
35 (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a

SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing  $10^7$  counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

5        The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography and compared.

#### 10    **VII.    Microarrays**

      A chemical coupling procedure and an ink jet device can be used to synthesize array elements on the surface of a substrate. (See, e.g., Baldeschweiler, supra.) An array analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced by hand  
15 or using available methods and machines and contain any appropriate number of elements. After hybridization, nonhybridized probes are removed and a scanner used to determine the levels and patterns of fluorescence. The degree of complementarity and the relative abundance of each probe which hybridizes to an element on the microarray may be assessed through analysis of the scanned images.

20        Full-length cDNAs, Expressed Sequence Tags (ESTs), or fragments thereof may comprise the elements of the microarray. Fragments suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). Full-length cDNAs, ESTs, or fragments thereof corresponding to one of the nucleotide sequences of the present invention, or selected at random from a cDNA library relevant to the present invention, are arranged on an  
25 appropriate substrate, e.g., a glass slide. The cDNA is fixed to the slide using, e.g., UV cross-linking followed by thermal and chemical treatments and subsequent drying. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645.) Fluorescent probes are prepared and used for hybridization to the elements on the substrate. The substrate is analyzed by procedures described above.

#### 30    **VIII.    Complementary Polynucleotides**

      Sequences complementary to the CYSKP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring CYSKP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are  
35 designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of CYSKP. To

inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the CYSKP-encoding transcript.

## 5 IX. Expression of CYSKP

Expression and purification of CYSKP is achieved using bacterial or virus-based expression systems. For expression of CYSKP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac (tac)* hybrid  
 10 promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express CYSKP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of CYSKP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus  
 15 (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding CYSKP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases.  
 20 Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E. K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, CYSKP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step,  
 25 affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from CYSKP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity  
 30 purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch 10 and 16). Purified CYSKP obtained by these methods can be used directly in the following activity assay.

## 35 X. Demonstration of CYSKP Activity

A microtubule motility assay for CYSKP activity measures motor domain function. In this assay, recombinant CYSKP is immobilized onto a glass slide or similar substrate. Taxol-stabilized bovine brain microtubules (commercially available) in a solution containing ATP and cytosolic extract are perfused onto the slide. Movement of microtubules as driven by CYSKP motor activity  
5 can be visualized and quantified using video-enhanced light microscopy and image analysis techniques. CYSKP activity is directly proportional to the frequency and velocity of microtubule movement.

In the alternative, an assay for CYSKP measures the binding affinity of CYSKP for actin as described by Hammell, R.L. and Hitchcock-DeGregori, S.E. (1997, J. Biol. Chem. 272:22409-  
10 22416). CYSKP and actin are prepared from in vitro recombinant cDNA expression systems and the N-terminus of CYSKP is acetylated using methods well known in the art. Binding of N-terminal acetyl-CYSKP to actin is measured by cosedimentation at 25°C in a Beckman model TL-100 centrifuge as described. The bound and free CYSKP are determined by quantitative densitometry of SDS-polyacrylamide gels stained with Coomassie Blue. Apparent binding constants ( $K_{app}$ ) and Hill  
15 coefficients (H) are determined by using methods well known in the art to fit the data to the equation as described by Hammell and Hitchcock-DeGregori (1997, supra). The CYSKP:actin ratio, determined using densitometry, is normalized. Hammell and Hitchcock-DeGregori (1997, supra) have shown that saturation of binding corresponds to a CYSKP:actin molar ratio of 0.14, a stoichiometry of 1 CYSKP:7 actin. The binding of CYSKP to actin is proportional to the CYSKP  
20 activity.

In the alternative, CYSKP are assayed by their ability to bind to F-actin using a blot overlay system similar to that described by Luna, E.J. et al. (1997, Soc. Gen. Physiol. Ser. 52:3-18). Proteins in plasma membrane-enriched cell extracts containing CYSKP are separated using SDS polyacrylamide gel electrophoresis (10% acrylamide). The gel-separated proteins are transferred to  
25 nitrocellulose using methods well known in the art and the blot is washed and pretreated with non-specific blocking agents. [ $^{125}$ I]-labeled F-actin is prepared and suspended in overlay buffer, then incubated with the blot for at least 16 hours at 4°C. Unbound label is washed with washing buffer, the blot is air dried and subjected to autoradiography for at least one hour. The autoradiograph band corresponding to the expected molecular mass of CYSKP is identified. The amount of observed  
30 [ $^{125}$ I]-labeled F-actin which binds to CYSKP is proportional to the amount of CYSKP present in the sample.

In the alternative, CYSKP activity is associated with its ability to form protein-protein complexes and is measured by its ability to regulate growth characteristics of NIH3T3 mouse fibroblast cells. A cDNA encoding CYSKP is subcloned into an appropriate eukaryotic expression  
35 vector. This vector is transfected into NIH3T3 cells using methods known in the art. Transfected



cells are compared with non-transfected cells for the following quantifiable properties: growth in culture to high density, reduced attachment of cells to the substrate, altered cell morphology, and ability to induce tumors when injected into immunodeficient mice. The activity of CYSKP is proportional to the extent of increased growth or frequency of altered cell morphology in NIH3T3  
5 cells transfected with CYSKP.

In the alternative, CYSKP activity is measured as ability to bind to microtubules. Microtubules are purified from adult rat brain by reversible assembly (Vallee, R. B. (1982) *Methods Enzymol.* 134:89-104) or the taxol method (Vallee, R. B. (1982) *J. Cell Biol.* 92:435-442) using PEM buffer (100 mM PIPES, pH 6.6, 1mM EGTA, 1mM MgSO<sub>4</sub>). To separate the MAPs from tubulin,  
10 the pellets from twice-cycled microtubules are resuspended in PEM buffer and applied to a 0.1 M MgSO<sub>4</sub>-saturated phosphocellulose column as described by Sloboda, R. D. and Rosenbaum, J. L. ((1982) *Methods Enzymol.* 85:409-416). The fractions containing protein are applied to a second phosphocellulose column. In a total volume of 100 ml, 20 ml of CYSKP (250 mg/ml) is added to 80 ml of whole microtubules (450 mg/ml) or tubulin (300 mg/ml) and incubated at 37 °C for 10 minutes  
15 in the presence of 1 mM GTP and 50 mM taxol. The suspension is centrifuged, the supernatant is removed, and the microtubule pellet is resuspended to the original reaction volume in PEM buffer. To assess the partitioning of CYSKP between the supernatant and pellet fractions, equal amounts of supernatant and resuspended pellet are placed in SDS sample buffer and assayed on a 5-20% gradient SDS polyacrylamide gel stained with Coomassie Brilliant Blue. The amount of CYSKP in the pellet  
20 fraction is proportional to the binding of CYSKP to microtubules.

In the alternative, CYSKP, or biologically active fragments thereof, are labeled with <sup>125</sup>I Bolton-Hunter reagent. (See, e.g., Bolton et al. (1973) *Biochem. J.* 133:529.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled CYSKP, washed, and any wells with labeled CYSKP complex are assayed. Data obtained using different  
25 concentrations of CYSKP are used to calculate values for the number, affinity, and association of CYSKP with the candidate molecules.

#### **XI. Functional Assays**

CYSKP function is assessed by expressing the sequences encoding CYSKP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a  
30 mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT (Life Technologies) and pCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 µg of recombinant vector are transiently transfected into a human cell line, preferably of endothelial or hematopoietic origin, using either liposome formulations or electroporation. 1-2 µg of an additional plasmid containing  
35 sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a

means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M. G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of CYSKP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding CYSKP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding CYSKP and other genes of interest can be analyzed by northern analysis or microarray techniques.

## **XII. Production of CYSKP Specific Antibodies**

CYSKP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the CYSKP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Perkin-Elmer) using fmoc-chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide activity by,

for example, binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

### **XIII. Purification of Naturally Occurring CYSKP Using Specific Antibodies**

Naturally occurring or recombinant CYSKP is substantially purified by immunoaffinity chromatography using antibodies specific for CYSKP. An immunoaffinity column is constructed by covalently coupling anti-CYSKP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing CYSKP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of CYSKP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/CYSKP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and CYSKP is collected.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
1	17	1285395	COLNNOT16	015834R1 (HUVLPB01), 866407T1 (BRAITUT03), 1232405F6 (LUNGFET03), 1285395H1 (COLNNOT16), 1478554T1 (CORPNOT02), 2103609R6 (BRAITUT02), 2254859R6 (OVRTUT01), 2692529H1 (LUNGNOT23), 2959263H1 (ADRENOT09), 3076303H2 (BONEUNT01), 3367129H1 (CONNTUT04), 3855643H1 (BRAITUT12), 4061729H1 (BRAINOT21), 4082537F6 (CONFNOT02)
2	18	1320252	BLADNOT04	229546R1 (PANCNOT01), 743845R6 (BRAITUT01), 826714T1 (PROSNOT06), 864534R1 (BRAITUT03), 997163R2 (KIDNTUT01), 1320252F6 and 1320252H1 (BLADNOT04), 1349551F1 (LATRTUT02), 1441011F1 (THYRNOT03), 1500649F6 (SINTBST01), 1525416T1 (UCMCL5T01), 1928370R6 (BRSTNOT02), 1932270H1 (COLNNOT16), 3213480F6 (BLADNOT08), 4540043H1 (THYRTMT01)
3	19	1259001	MENITUT03	1259001H1 (MENITUT03), 1550766H1 (PROSNOT06), 1594658F1 (BRAINOT14), 1594658T1 (BRAINOT14), 1653882F6 (PROSTUT08), 1864111F6 (PROSNOT19), 3399605H1 (UTRSNOT16), 3677286H1 (PLACNOT07), 5045012H1 (PLACFER01), 5188326H1 (LUNGTMOT04), SATA00218F1, SATA00850F1
4	20	1627027	COLNPOT01	1361332F6 (LUNGNOT12), 1933148H1 (COLNNOT16), 2378239F6 (ISLTNOT01), 2378239T6 (ISLTNOT01), 3433415H1 (PENCNOT05), 3433415X303F1 (PENCNOT05), 4453336H1 (HEAADIR01)
5	21	1905315	OVARNOT07	1504617F1 (BRAITUT07), 1520641F1 (BLADTUT04), 1905315H1 (OVARNOT07), 3282914F6 (HEAONOT05), 3282914T6 (HEAONOT05)
6	22	1997789	BRSTTUT03	833978T1 (PROSNOT07), 1309235R1 (COLNFET02), 1659579F6 (URETTUT01), 1734634T6 (COLNNOT22), 2930134F6 (TLYMNOT04), SAEA00063F1
7	23	2303465	BRSTNOT05	411540R6 (BRSTNOT01), 487448F1 (HNT2AGT01), 487448R1 (HNT2AGT01), 647107H1 (BRSTTUT02), 1426319F1 (SINTBST01), 2155735F6 (BRAINOT09), 2155735T6 (BRAINOT09), 2303465H1 (BRSTNOT05)

Table 1 (cont.)

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
8	24	2363178	LUNGFET05	2363178H1 (LUNGFET05), 2590354F6 (LUNGN0T22), 2590354T6 (LUNGN0T22)
9	25	2363327	ADREN0T07	013068R6 (THP1PLB01), 1295235H1 (PGANN0T03), 1445845X13 (PLACN0T02), 1472260R6 (LUNGTUT03), 1474238T6 (LUNGTUT03), 1643970F6 (HEARFET01), 1794319R6 (PROSTUT05), 1868517F6 (SKINBIT01), 2057830R6 (BEPIN0T01), 2058164H1 (BEPIN0T01), 2363327F6 (ADREN0T07), 2363327H1 (ADREN0T07), 2363327T6 (ADREN0T07), 2877024F6 (THYRN0T10), 2877024T6 (THYRN0T10), 2930751F6 (TLYMN0T04), 3002267F6 (TLYMN0T06)
10	26	2508327	CONUTUT01	2508327H1 (CONUTUT01), 2508327T6 (CONUTUT01), 3743046H1 (THYMN0T08)
11	27	2524555	BRAITUT21	781951H1 (MYOMN0T01), 2524555H1 (BRAITUT21), 3243902H1 (BRAIN0T19), 4296903H1 (SCOMDIT01), SAEA01358F1
12	28	2900717	DRGCN0T01	933857R1 (CERVN0T01), 1632793T6 (COLNN0T19), 1909014F6 (CONNTUT01), 2250618R6 (OVARUT01), 2900717F6 (DRGCN0T01), 2900717H1 (DRGCN0T01), 2967545H1 (SCORN0T04), 3506152H1 (ADREN0T11), 4713710H1 (BRAIHCT01)
13	29	3088904	HEAON0T03	990189H1 (COLNN0T11), 2530228H1 (GBLAN0T02), 3088904F6 (HEAON0T03), 3088904H1 (HEAON0T03), 3176845T6 (UTRSTUT04)
14	30	3745193	THYMN0T08	2775454H1 (PANCN0T15), 2811439H1 (OVARN0T10), 3745193F6 (THYMN0T08), 3745193H1 (THYMN0T08), 3745193T6 (THYMN0T08)
15	31	3822123	BONSTUT01	736132R6 (TONSN0T01), 1856649F6 (PROSN0T18), 1877413F6 (LEUKN0T03), 3395569T6 (LUNGN0T28), 3577567H1 (BRONN0T01), 3822123H1 (BONSTUT01), 4247960H1 (BRABDIT01)

Table 1 (cont.)

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
16	32	4217506	ADRENOT15	590362R1 (UTRSNOT01), 973313R6 (MUSCNOT02), 4216992H1 (ADRENOT15), SBJA03360F1

Table 2

Protein SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences	Homologous Sequences	Analytical Methods
1	1005	S15 T194 S402 S548 S614 T673 S7 S39 S151 S159 T180 T223 T530 S647 S658 S682 T730 S744 S746 S748 T766 S828 S854 T879 S890 T952 S58 S208 T212 S323 T381 S449 S518 S543 S544 S884 T944 Y623	N36	L225 to F263 W272 to D300 I321 to Q367 H379 to L408 F440 to K458 S718 to G721	protein 4.1	BLAST ProfileScan PFAM BLOCKS PRINTS
2	1045	T92 S270 S366 S23 T150 T207 T396 S418 T448 T525 S549 S571 S706 T811 S815 S840 S842 S872 S878 T883 S889 T898 S923 S966 S987 S1038 S36 S41 S336 S340 T343 S370 S408 T538 T551 S657 S658 S770 T789 T826 S839 Y542	N152 N495 N919	L47 to F85 W94 to D123 L144 to D190 I196 to I249 F261 to K279 S770 to G773	protein 4.1	BLAST ProfileScan PFAM BLOCKS PRINTS

Table 2 (cont.)

Protein SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences	Homologous Sequences	Analytical Methods
3	324	T71 T78 S121 S123 S225 T259 S283 T304 S144 S150 S181 T249 S273 S276 S289	N151 N180	Ankyrin repeat: P6-A41; D42-E74; G76-K85	ARF-directed GTPase activating protein (ankyrin-repeat containing, involved in regulation of cytoskeletal organization) [Mus musculus] g4063614	BLAST PFAM BLOCKS_PFAM
4	385	T6 T30 T375 T19 S61 S161 S176	N59 N132 N328 N341		cardiac muscle tensin [Gallus gallus] g619577	BLAST
5	364	T65 T74 T79 S80 T139 T151 T228 T244 T276 T9 T79 T349 Y116	N318		similar to alpha- actinin [Caenorhabditis elegans] g2315828	BLAST
6	395	S158 S94 S130 S213 S214 S251 S283 S296 S348 T19 S20 T46 S121 T250 S285	N62 N317	WASP homology domain 1: M1-L109	ena-VASP like protein [Mus musculus] g1644453	PFAM BLAST



Table 2 (cont.)

Protein SEQ ID No:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences	Homologous Sequences	Analytical Methods
7	523	T213 S140 T157 S215 T245 S251 S286 S516 T518 S57 T342 S398 S405 S427 S453 S483 T484 T503 Y103 Y197 Y297	N410	ATP/GTP-binding site motif A (P loop): G74-T82	dynein light chain A [Gallus sp.] g510249	MOTIFS BLAST
8	348	S32 S55 T104 T153 T183 S213 T223 T249 S34 S41 T51 T52 S166 S293	N86 N164 N233	Tektin signature: R119-E139	tektin C1 [Strongylocentrotus purpuratus] g1353490	PRINTS BLAST
9	731	S117 S136 S162 T168 S219 S249 S390 T451 S665 S694 S15 T292 S313 T559 S703 Y131 Y407 Y490	N125 N134 N205 N551	CAP-Gly domain proteins: G40-C64		BLOCKS
10	147	T100 S137 S138 S9 Y86 Y116	N91		light chain 3 subunit of microtubule- associated proteins 1A and 1B [Rattus norvegicus] g455109	BLAST

Table 2 (cont.)

Protein SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences	Homologous Sequences	Analytical Methods
11	57	T13 S15		Thymosin beta-4 family: S15-G55	thymosin beta-4 [Mus musculus] g54794	PFAM MOTIFS ProfileScan BLOCKS BLAST
12	452	T8 S36 S75 T94 S117 S237 S246 S311 T358 S406 T4 S208 S216 T239 S295 Y188	N21 N80 N91 N373	ATP/GTP-binding site motif A (P loop): G204-S211  Signal peptide: M1-G34	non-A non-B hepatitis- associated microtubular aggregates protein (p44) [Pan trogloodytes] g218576	MOTIFS SPSCAN BLAST
13	281	T76 T50 S60 S207 S212 T213 T234 T249 S259 T274 S120 S155		Tropomyosins: K6-E38; K45-L281	beta-tropomyosin [Mus musculus] g192157	MOTIFS PFAM ProfileScan BLOCKS BLAST
14	92	S16 S23 T45 T60 T71 T85 T78		Tropomyosin: M1-M92	tropomyosin 5 TM-5 [Rattus sp.] g1703676	PFAM ProfileScan BLOCKS BLAST

Table 2 (cont.)

Protein SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences	Homologous Sequences	Analytical Methods
15	448	T126 T73 T94 S165 T193 S287 S439 T82 S241 T337 Y172	N380	Tubulin: M1-E433  Signal peptide: M1-A32	alpha-tubulin isotype M-alpha-6 [Mus musculus] g202215	MOTIFS PFAM BLOCKS SPSCAN
16	269	T181 S2 T40 S88 S244 T241 S253	N164	Troponin: K73-W215; H252-K269	troponin T fast muscle isoform [Mus musculus] g2340062	PFAM BLOCKS PFAM BLAST

Table 3

Nucleotide SEQ ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
17	549-587	Nervous (0.265) Reproductive (0.229) Cardiovascular (0.145)	Cancer (0.482) Inflammation (0.253)	pINCY
18	882-918	Reproductive (0.220) Nervous (0.207) Gastrointestinal (0.134)	Cancer (0.549) Trauma (0.110) Inflammation (0.098)	pINCY
19	817-864	Reproductive (0.372) Nervous (0.186) Gastrointestinal (0.116)	Cancer and Cell Proliferation (0.651) Inflammation and Immune Response (0.256)	pINCY
20	489-533	Gastrointestinal (0.385) Cardiovascular (0.154) Reproductive (0.154)	Cancer and Cell Proliferation (0.385) Inflammation and Immune Response (0.308)	pINCY
21	50-106	Reproductive (0.220) Hematopoietic/Immune (0.200) Cardiovascular (0.140)	Cancer and Cell Proliferation (0.500) Inflammation and Immune Response (0.360)	pINCY
22	1070-1228	Hematopoietic/Immune (0.211) Reproductive (0.186) Nervous (0.180)	Cancer and Cell Proliferation (0.590) Inflammation and Immune Response (0.360)	PSPORT1
23	250-336	Reproductive (0.291) Gastrointestinal (0.163) Cardiovascular (0.116) Nervous (0.116)	Cancer and Cell Proliferation (0.663) Inflammation and Immune Response (0.337)	PSPORT1

Table 3 (cont.)

Nucleotide SEQ ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
24	164-208	Cardiovascular (0.333) Developmental (0.333) Reproductive (0.333)	Inflammation and Immune Response (0.500) Cancer and Cell Proliferation (0.500)	PSPORT1
25	1028-1072	Hematopoietic/Immune (0.286) Reproductive (0.159) Nervous (0.127)	Cancer and Cell Proliferation (0.540) Inflammation and Immune Response (0.413)	pINCY
26	397-516	Gastrointestinal (0.333) Hematopoietic/Immune (0.333) Musculoskeletal (0.333)	Cancer and Cell Proliferation (0.667)	pINCY
27	434-541	Reproductive (0.236) Nervous (0.156) Gastrointestinal (0.148)	Cancer and Cell Proliferation (0.575) Inflammation and Immune Response (0.353)	pINCY
28	1-177	Reproductive (0.269) Hematopoietic/Immune (0.192) Nervous (0.192)	Cancer and Cell Proliferation (0.654) Inflammation and Immune Response (0.462)	pINCY
29		Reproductive (0.339) Gastrointestinal (0.191) Cardiovascular (0.114)	Cancer and Cell Proliferation (0.631) Inflammation and Immune Response (0.288)	pINCY
30	488-532 551-649	Reproductive (0.199) Gastrointestinal (0.144) Nervous (0.144)	Cancer and Cell Proliferation (0.580) Inflammation and Immune Response (0.326)	pINCY

Table 3 (cont.)

Nucleotide SEQ ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
31	163-207	Nervous (0.305) Reproductive (0.158) Gastrointestinal (0.137)	Cancer and Cell Proliferation (0.547) Inflammation and Immune Response (0.295)	pINCY
32	99-143	Musculoskeletal (0.280) Developmental (0.160) Reproductive (0.160)	Cancer and Cell Proliferation (0.680) Inflammation and Immune Response (0.280)	pINCY

Table 4

Polynucleotide SEQ ID NO:	Library	Library Comment
17	COLNNOT16	Library was constructed using RNA isolated from sigmoid colon tissue removed from a 62-year-old Caucasian male during a sigmoidectomy and permanent colectomy. Pathology for the associated tumor tissue indicated invasive grade 2 adenocarcinoma, with invasion through the muscularis. One lymph node contained metastasis with extranodal extension. Family history included benign hypertension, atherosclerotic coronary artery disease, breast cancer, and prostate cancer.
18	BLADNOT04	Library was constructed using RNA isolated from bladder tissue of a 28-year-old Caucasian male, who died from a self-inflicted gunshot wound. The patient had a history of alcohol and tobacco use.
19	MENITUT03	Library was constructed using RNA isolated from brain meningioma tissue removed from a 35-year-old Caucasian female during excision of a cerebral meningeal lesion. Pathology indicated a benign neoplasm in the right cerebellopontine angle of the brain. Patient history included hypothyroidism. Family history included myocardial infarction and breast cancer.
20	COLNPOT01	Library was constructed using RNA isolated from colon polyp tissue removed from a 40-year-old Caucasian female during a total colectomy. Pathology indicated an inflammatory pseudopolyp; this tissue was associated with a focally invasive grade 2 adenocarcinoma and multiple tubovillous adenomas. Patient history included a benign neoplasm of the bowel.
21	OVARNOT07	Library was constructed using RNA isolated from left ovarian tissue removed from a 28-year-old Caucasian female during a vaginal hysterectomy and removal of the fallopian tubes and ovaries. The tissue was associated with multiple follicular cysts, endometrium in a weakly proliferative phase, and chronic cervicitis of the cervix with squamous metaplasia. Family history included benign hypertension, hyperlipidemia, and atherosclerotic coronary artery disease.

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Library	Library Comment
22	BRSTTUT03	Library was constructed using RNA isolated from breast tumor tissue removed from a 58-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated multicentric invasive grade 4 lobular carcinoma. The mass was identified in the upper outer quadrant, and three separate nodules were found in the lower outer quadrant of the left breast. Patient history included skin cancer, rheumatic heart disease, osteoarthritis, and tuberculosis. Family history included cerebrovascular disease, coronary artery aneurysm, breast cancer, prostate cancer, atherosclerotic coronary artery disease, and type I diabetes.
23	BRSTNOT05	Library was constructed using RNA isolated from breast tissue removed from a 58-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology for the associated tumor tissue indicated multicentric invasive grade 4 lobular carcinoma. Patient history included skin cancer, rheumatic heart disease, osteoarthritis, and tuberculosis. Family history included cerebrovascular and cardiovascular disease, breast and prostate cancer, and type I diabetes.
24	LUNGFET05	Library was constructed using RNA isolated from lung tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation from anencephalus.
25	ADRENOT07	Library was constructed using RNA isolated from adrenal tissue removed from a 61-year-old female during a bilateral adrenalectomy. Patient history included an unspecified disorder of the adrenal glands.
26	CONUTUT01	Library was constructed using RNA isolated from sigmoid mesentery tumor tissue obtained from a 61-year-old female during a total abdominal hysterectomy and bilateral salpingo-oophorectomy with regional lymph node excision. Pathology indicated a metastatic grade 4 malignant mixed müllerian tumor present in the sigmoid mesentery at two sites.



Table 4 (cont.)

Polynucleotide SEQ ID NO:	Library	Library Comment
27	BRAITUT21	Library was constructed using RNA isolated from brain tumor tissue removed from the midline frontal lobe of a 61-year-old Caucasian female during excision of a cerebral meningeal lesion. Pathology indicated subfrontal meningothelial meningioma with no atypia. One ethmoid and mucosal tissue sample indicated meningioma. Family history included cerebrovascular disease, senile dementia, hyperlipidemia, benign hypertension, atherosclerotic coronary artery disease, congestive heart failure, and breast cancer.
28	DRGCNOT01	Library was constructed using RNA isolated from dorsal root ganglion tissue removed from the cervical spine of a 32-year-old Caucasian male who died from acute pulmonary edema and bronchopneumonia, bilateral pleural and pericardial effusions, and malignant lymphoma (natural killer cell type). Patient history included probable cytomegalovirus infection, hepatic congestion and steatosis, splenomegaly, hemorrhagic cystitis, thyroid hemorrhage, and Bell's palsy. Surgeries included colonoscopy, large intestine biopsy, adenotonsillectomy, and nasopharyngeal endoscopy and biopsy; treatment included radiation therapy.
29	HEAONOT03	Library was constructed using RNA isolated from aortic tissue removed from a 27-year-old Caucasian female, who died from an intracranial bleed.
30	THYMNOT08	Library was constructed using RNA isolated from thymus tissue removed from a 4-month-old Caucasian male during a total thymectomy and open heart repair of atrioventricular canal defect using hypothermia. The patient presented with a congenital heart anomaly, congestive heart failure, and Down syndrome. Patient history included abnormal thyroid function study and premature birth. Previous procedures included right and left heart angiocardiography.
31	BONSTUT01	Library was constructed using RNA isolated from sacral bone tumor tissue removed from an 18-year-old Caucasian female during an exploratory laparotomy with soft tissue excision. Pathology indicated giant cell tumor of the sacrum. Patient history included a soft tissue malignant neoplasm. Family history included prostate cancer.

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Library	Library Comment
32	ADRENOT15	Library was constructed using RNA isolated from adrenal tissue removed from a Caucasian female fetus, who died from anencephalus after 16-weeks' gestation.

Table 5

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25: 3389-3402.	ESTs: Probability value= 1.0E-8 or less Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183: 63-98; and Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.0E-6 Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less Full Length sequences: fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS and PRINTS databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff, Nucl. Acid Res., 19:6565-72, 1991. J.G. Henikoff and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37: 417-424.	Score=1000 or greater; Ratio of Score/Strength = 0.75 or larger; and Probability value= 1.0E-3 or less if applicable
PFAM	A Hidden Markov Models-based application useful for protein family search.	Krogh, A. et al. (1994) J. Mol. Biol., 235:1501-1531; Somnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322.	Score=10-50 bits, depending on individual protein families

Table 5

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25: 3389-3402.	ESTs: Probability value= 1.0E-8 or less Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183: 63-98; and Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.0E-6 Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less Full Length sequences: fastx score= 100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS and PRINTS databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff, Nucl. Acid Res., 19:6565-72, 1991. J.G. Henikoff and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37: 417-424.	Score=1000 or greater; Ratio of Score/Strength = 0.75 or larger; and Probability value= 1.0E-3 or less if applicable
PFAM	A Hidden Markov Models-based application useful for protein family search.	Krogh, A. et al. (1994) J. Mol. Biol., 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322.	Score=10-50 bits, depending on individual protein families

What is claimed is:

1. A substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, and fragments thereof.
2. A substantially purified variant having at least 90% amino acid sequence identity to the amino acid sequence of claim 1.
3. An isolated and purified polynucleotide encoding the polypeptide of claim 1.
4. An isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide of claim 3.
5. An isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide of claim 3.
6. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide of claim 3.
7. A method for detecting a polynucleotide, the method comprising the steps of:
  - (a) hybridizing the polynucleotide of claim 6 to at least one nucleic acid in a sample, thereby forming a hybridization complex; and
  - (b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of the polynucleotide in the sample.
8. The method of claim 7 further comprising amplifying the polynucleotide prior to hybridization.
9. An isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, and fragments thereof.

10. An isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide of claim 9.
11. An isolated and purified polynucleotide having a sequence which is complementary  
5 to the polynucleotide of claim 9.
12. An expression vector comprising at least a fragment of the polynucleotide of claim 3.
13. A host cell comprising the expression vector of claim 12.
- 10 14. A method for producing a polypeptide, the method comprising the steps of:  
a) culturing the host cell of claim 13 under conditions suitable for the expression of the polypeptide; and  
b) recovering the polypeptide from the host cell culture.
- 15 15. A pharmaceutical composition comprising the polypeptide of claim 1 in conjunction with a suitable pharmaceutical carrier.
16. A purified antibody which specifically binds to the polypeptide of claim 1.
- 20 17. A purified agonist of the polypeptide of claim 1.
18. A purified antagonist of the polypeptide of claim 1.
- 25 19. A method for treating or preventing a disorder associated with decreased expression or activity of CYSKP, the method comprising administering to a subject in need of such treatment an effective amount of the pharmaceutical composition of claim 15.
- 30 20. A method for treating or preventing a disorder associated with increased expression or activity of CYSKP, the method comprising administering to a subject in need of such treatment an effective amount of the antagonist of claim 18.

## SEQUENCE LISTING

<110> INCYTE PHARMACEUTICALS, INC.

LAL, Preeti

TANG, Y. Tom

YUE, Henry

HILLMAN, Jennifer L.

BANDMAN, Olga

CORLEY, Neil C.

GUEGLER, Karl J.

PATTERSON, Chandra

AZIMZAI, Yalda

BAUGHN, Mariah R.

## <120> HUMAN CYTOSKELETON ASSOCIATED PROTEINS

<130> PF-0594 PCT

<140> To Be Assigned

<141> Herewith

<150> 09/156,470; unassigned; 60/131,321

<151> 1998-09-18; 1998-09-18; 1999-04-27

<160> 32

<170> PERL Program

<210> 1

<211> 1005

<212> PRT

<213> Homo sapiens

<220>

```
<221> misc_feature
```

<223> Incyte ID No: 1285395

<400> 1

Met	Thr	Thr	Glu	Val	Gly	Ser	Val	Ser	Glu	Val	Lys	Lys	Asp	Ser
1				5					10					15
Ser	Gln	Leu	Gly	Thr	Asp	Ala	Thr	Lys	Glu	Lys	Pro	Lys	Glu	Val
				20					25					30
Ala	Glu	Asn	Gln	Gln	Asn	Gln	Ser	Ser	Asp	Pro	Glu	Glu	Glu	Lys
				35					40					45
Gly	Ser	Gln	Pro	Pro	Pro	Ala	Ala	Glu	Ser	Gln	Ser	Ser	Leu	Arg
				50					55					60
Arg	Gln	Lys	Arg	Glu	Lys	Glu	Thr	Ser	Glu	Ser	Arg	Gly	Ile	Ser
				65					70					75
Arg	Phe	Ile	Pro	Pro	Trp	Leu	Lys	Lys	Gln	Lys	Ser	Tyr	Thr	Leu
				80					85					90
Val	Val	Ala	Lys	Asp	Gly	Gly	Asp	Lys	Lys	Glu	Pro	Thr	Gln	Ala
				95					100					105
Val	Val	Glu	Glu	Gln	Val	Leu	Asp	Lys	Glu	Glu	Pro	Leu	Pro	Glu
				110					115					120

Glu	Gln	Arg	Gln	Ala	Lys	Gly	Asp	Ala	Glu	Glu	Met	Ala	Gln	Lys	125	130	135
Lys	Gln	Glu	Ile	Lys	Val	Glu	Val	Lys	Glu	Glu	Lys	Pro	Ser	Val	140	145	150
Ser	Lys	Glu	Glu	Lys	Pro	Ser	Val	Ser	Lys	Val	Glu	Met	Gln	Pro	155	160	165
Thr	Glu	Leu	Val	Ser	Lys	Glu	Arg	Glu	Glu	Lys	Val	Lys	Glu	Thr	170	175	180
Gln	Glu	Asp	Lys	Leu	Glu	Gly	Gly	Ala	Ala	Lys	Arg	Glu	Thr	Lys	185	190	195
Glu	Val	Gln	Thr	Asn	Glu	Leu	Lys	Ala	Glu	Lys	Ala	Ser	Gln	Lys	200	205	210
Val	Thr	Lys	Lys	Thr	Lys	Thr	Val	Gln	Cys	Lys	Val	Thr	Leu	Leu	215	220	225
Asp	Gly	Thr	Glu	Tyr	Ser	Cys	Asp	Leu	Glu	Lys	His	Ala	Lys	Gly	230	235	240
Gln	Val	Leu	Phe	Asp	Lys	Val	Cys	Glu	His	Leu	Asn	Leu	Leu	Glu	245	250	255
Lys	Asp	Tyr	Phe	Gly	Leu	Leu	Phe	Gln	Glu	Ser	Pro	Glu	Gln	Lys	260	265	270
Asn	Trp	Leu	Asp	Pro	Ala	Lys	Glu	Ile	Lys	Arg	Gln	Leu	Arg	Asn	275	280	285
Leu	Pro	Trp	Leu	Phe	Thr	Phe	Asn	Val	Lys	Phe	Tyr	Pro	Pro	Asp	290	295	300
Pro	Ser	Gln	Leu	Thr	Glu	Asp	Ile	Thr	Arg	Tyr	Phe	Leu	Cys	Leu	305	310	315
Gln	Leu	Arg	Gln	Asp	Ile	Ala	Ser	Gly	Arg	Leu	Pro	Cys	Ser	Phe	320	325	330
Val	Thr	His	Ala	Leu	Leu	Gly	Ser	Tyr	Thr	Leu	Gln	Ala	Glu	Leu	335	340	345
Gly	Asp	Tyr	Asp	Pro	Glu	Glu	His	Gly	Ser	Ile	Asp	Leu	Ser	Glu	350	355	360
Phe	Gln	Phe	Ala	Pro	Thr	Gln	Thr	Lys	Glu	Leu	Glu	Glu	Lys	Val	365	370	375
Ala	Glu	Leu	His	Lys	Thr	His	Arg	Gly	Leu	Ser	Pro	Ala	Gln	Ala	380	385	390
Asp	Ser	Gln	Phe	Leu	Glu	Asn	Ala	Lys	Arg	Leu	Ser	Met	Tyr	Gly	395	400	405
Val	Asp	Leu	His	His	Ala	Lys	Asp	Ser	Glu	Gly	Val	Asp	Ile	Lys	410	415	420
Leu	Gly	Val	Cys	Ala	Asn	Gly	Leu	Leu	Ile	Tyr	Lys	Asp	Arg	Leu	425	430	435
Arg	Ile	Asn	Arg	Phe	Ala	Trp	Pro	Lys	Ile	Leu	Lys	Ile	Ser	Tyr	440	445	450
Lys	Arg	Ser	Asn	Phe	Tyr	Ile	Lys	Val	Arg	Pro	Ala	Glu	Leu	Glu	455	460	465
Gln	Phe	Glu	Ser	Thr	Ile	Gly	Phe	Lys	Leu	Pro	Asn	His	Arg	Ala	470	475	480
Ala	Lys	Arg	Leu	Trp	Lys	Val	Cys	Val	Glu	His	His	Thr	Phe	Tyr	485	490	495
Arg	Leu	Val	Ser	Pro	Glu	Gln	Pro	Pro	Lys	Ala	Lys	Phe	Leu	Thr	500	505	510
Leu	Gly	Ser	Lys	Phe	Arg	Tyr	Ser	Gly	Arg	Thr	Gln	Ala	Gln	Thr	515	520	525
Arg	Gln	Ala	Ser	Thr	Leu	Ile	Asp	Arg	Pro	Ala	Pro	His	Phe	Glu			



	530		535		540
Arg Thr Ser Ser	Lys Arg Val Ser Arg	Ser Leu Asp Gly Ala	Pro		
	545		550		555
Ile Gly Val Met	Asp Gln Ser Leu Met	Lys Asp Phe Pro Gly	Ala		
	560		565		570
Ala Gly Glu Ile	Ser Ala Tyr Gly Pro	Gly Leu Val Ser Ile	Ala		
	575		580		585
Val Val Gln Asp	Gly Asp Gly Arg Arg	Glu Val Arg Ser Pro	Thr		
	590		595		600
Lys Ala Pro His	Leu Gln Leu Ile Glu	Gly Lys Lys Asn Ser	Leu		
	605		610		615
Arg Val Glu Gly	Asp Asn Ile Tyr Val	Arg His Ser Asn Leu	Met		
	620		625		630
Leu Glu Glu Leu	Asp Lys Ala Gln Glu	Asp Ile Leu Lys His	Gln		
	635		640		645
Ala Ser Ile Ser	Glu Leu Lys Arg Asn	Phe Met Glu Ser Thr	Pro		
	650		655		660
Glu Pro Arg Pro	Asn Glu Trp Glu Lys	Arg Arg Ile Thr Pro	Leu		
	665		670		675
Ser Leu Gln Thr	Gln Gly Ser Ser His	Glu Thr Leu Asn Ile	Val		
	680		685		690
Glu Glu Lys Lys	Arg Ala Glu Val Gly	Lys Asp Glu Arg Val	Ile		
	695		700		705
Thr Glu Glu Met	Asn Gly Lys Glu Ile	Ser Pro Gly Ser Gly	Pro		
	710		715		720
Gly Glu Ile Arg	Lys Val Glu Pro Val	Thr Gln Lys Asp Ser	Thr		
	725		730		735
Ser Leu Ser Ser	Glu Ser Ser Ser Ser	Ser Ser Glu Ser Glu	Glu		
	740		745		750
Glu Asp Val Gly	Glu Tyr Arg Pro His	His Arg Val Thr Glu	Gly		
	755		760		765
Thr Ile Arg Glu	Glu Gln Glu Tyr Glu	Glu Glu Val Glu Glu	Glu		
	770		775		780
Pro Arg Pro Ala	Ala Lys Val Val Glu	Arg Glu Glu Ala Val	Pro		
	785		790		795
Glu Ala Ser Pro	Val Thr Gln Ala Gly	Ala Ser Val Ile Thr	Val		
	800		805		810
Glu Thr Val Ile	Gln Glu Asn Val Gly	Ala Gln Lys Ile Pro	Gly		
	815		820		825
Glu Lys Ser Val	His Glu Gly Ala Leu	Lys Gln Asp Met Gly	Glu		
	830		835		840
Glu Ala Glu Glu	Glu Pro Gln Lys Val	Asn Gly Glu Val Ser	His		
	845		850		855
Val Asp Ile Asp	Val Leu Pro Gln Ile	Ile Cys Cys Ser Glu	Pro		
	860		865		870
Pro Val Val Lys	Thr Glu Met Val Thr	Ile Ser Asp Ala Ser	Gln		
	875		880		885
Arg Thr Glu Ile	Ser Thr Lys Glu Val	Pro Ile Val Gln Thr	Glu		
	890		895		900
Thr Lys Thr Ile	Thr Tyr Glu Ser Pro	Gln Ile Asp Gly Gly	Ala		
	905		910		915
Gly Gly Asp Ser	Gly Thr Leu Leu Thr	Ala Gln Thr Ile Thr	Ser		
	920		925		930
Glu Ser Val Ser	Thr Thr Thr Thr Thr	His Ile Thr Lys Thr	Val		
	935		940		945

```

Lys Gly Gly Ile Ser Glu Thr Arg Ile Glu Lys Arg Ile Val Ile
          950          955          960
Thr Gly Asp Gly Asp Ile Asp His Asp Gln Ala Leu Ala Gln Ala
          965          970          975
Ile Arg Glu Ala Arg Glu Gln His Pro Asp Met Ser Val Thr Arg
          980          985          990
Val Val Val His Lys Glu Thr Glu Leu Ala Glu Glu Gly Glu Asp
          995          1000          1005

```

<210> 2

<211> 1045

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 1320252

<400> 2

```

Met Gly Glu Ile Glu Gln Arg Pro Thr Pro Gly Ser Arg Leu Gly
  1          5          10          15
Ala Pro Glu Asn Ser Gly Ile Ser Thr Leu Glu Arg Gly Gln Lys
          20          25          30
Pro Pro Pro Thr Pro Ser Gly Lys Leu Val Ser Ile Lys Ile Gln
          35          40          45
Met Leu Asp Asp Thr Gln Glu Ala Phe Glu Val Pro Gln Arg Ala
          50          55          60
Pro Gly Lys Val Leu Leu Asp Ala Val Cys Asn His Leu Asn Leu
          65          70          75
Val Glu Gly Asp Tyr Phe Gly Leu Glu Phe Pro Asp His Lys Lys
          80          85          90
Ile Thr Val Trp Leu Asp Leu Leu Lys Pro Ile Val Lys Gln Ile
          95          100          105
Arg Arg Pro Lys His Val Val Val Lys Phe Val Val Lys Phe Phe
          110          115          120
Pro Pro Asp His Thr Gln Leu Gln Glu Glu Leu Thr Arg Tyr Leu
          125          130          135
Phe Ala Leu Gln Val Lys Gln Asp Leu Ala Gln Gly Arg Leu Thr
          140          145          150
Cys Asn Asp Thr Ser Ala Ala Leu Leu Ile Ser His Ile Val Gln
          155          160          165
Ser Glu Ile Gly Asp Phe Asp Glu Ala Leu Asp Arg Glu His Leu
          170          175          180
Ala Lys Asn Lys Tyr Ile Pro Gln Gln Asp Ala Leu Glu Asp Lys
          185          190          195
Ile Val Glu Phe His His Asn His Ile Gly Gln Thr Pro Ala Glu
          200          205          210
Ser Asp Phe Gln Leu Leu Glu Ile Ala Arg Arg Leu Glu Met Tyr
          215          220          225
Gly Ile Arg Leu His Pro Ala Lys Asp Arg Glu Gly Thr Lys Ile
          230          235          240
Asn Leu Ala Val Ala Asn Thr Gly Ile Leu Val Phe Gln Gly Phe
          245          250          255
Thr Lys Ile Asn Ala Phe Asn Trp Ala Lys Val Arg Lys Leu Ser

```

	260		265		270
Phe Lys Arg Lys Arg Phe Phe Ile Lys		Leu Arg Pro Asp Ala Asn			
	275		280		285
Ser Ala Tyr Gln Asp Thr Leu Glu Phe		Leu Met Ala Ser Arg Asp			
	290		295		300
Phe Cys Lys Ser Phe Trp Lys Ile Cys		Val Glu His His Ala Phe			
	305		310		315
Phe Arg Leu Phe Glu Glu Pro Lys Pro		Lys Pro Lys Pro Val Leu			
	320		325		330
Phe Ser Arg Gly Ser Ser Phe Arg Phe		Ser Gly Arg Thr Gln Lys			
	335		340		345
Gln Val Leu Asp Tyr Val Lys Glu Gly		Gly His Lys Lys Val Gln			
	350		355		360
Phe Glu Arg Lys His Ser Lys Ile His		Ser Ile Arg Ser Leu Ala			
	365		370		375
Ser Gln Pro Thr Glu Leu Asn Ser Glu		Val Leu Glu Gln Ser Gln			
	380		385		390
Gln Ser Thr Ser Leu Thr Phe Gly Glu		Gly Ala Glu Ser Pro Gly			
	395		400		405
Gly Gln Ser Cys Arg Arg Gly Lys Glu		Pro Lys Val Ser Ala Gly			
	410		415		420
Glu Pro Gly Ser His Pro Ser Pro Ala		Pro Arg Arg Ser Pro Ala			
	425		430		435
Gly Asn Lys Gln Ala Asp Gly Ala Ala		Ser Ala Pro Thr Glu Glu			
	440		445		450
Glu Glu Glu Val Val Lys Asp Arg Thr		Gln Gln Ser Lys Pro Gln			
	455		460		465
Pro Pro Gln Pro Ser Thr Gly Ser Leu		Thr Gly Ser Pro His Leu			
	470		475		480
Ser Glu Leu Ser Val Asn Ser Gln Gly		Gly Val Ala Pro Ala Asn			
	485		490		495
Val Thr Leu Ser Pro Asn Leu Ser Pro		Asp Thr Lys Gln Ala Ser			
	500		505		510
Pro Leu Ile Ser Pro Leu Leu Asn Asp		Gln Ala Cys Pro Arg Thr			
	515		520		525
Asp Asp Glu Asp Glu Gly Arg Arg Lys		Arg Phe Pro Thr Asp Lys			
	530		535		540
Ala Tyr Phe Ile Ala Lys Glu Val Ser		Thr Thr Glu Arg Thr Tyr			
	545		550		555
Leu Lys Asp Leu Glu Val Ile Thr Ser		Trp Phe Gln Ser Thr Val			
	560		565		570
Ser Lys Glu Asp Ala Met Pro Glu Ala		Leu Lys Ser Leu Ile Phe			
	575		580		585
Pro Asn Phe Glu Pro Leu His Lys Phe		His Thr Asn Phe Leu Lys			
	590		595		600
Glu Ile Glu Gln Arg Leu Ala Leu Trp		Glu Gly Arg Ser Asn Ala			
	605		610		615
Gln Ile Arg Asp Tyr Gln Arg Ile Gly		Asp Val Met Leu Lys Asn			
	620		625		630
Ile Gln Gly Met Lys His Leu Ala Ala		His Leu Trp Lys His Ser			
	635		640		645
Glu Ala Leu Glu Ala Leu Glu Asn Gly		Ile Lys Ser Ser Arg Arg			
	650		655		660
Leu Glu Asn Phe Cys Arg Asp Phe Glu		Leu Gln Lys Val Cys Tyr			
	665		670		675

Leu	Pro	Leu	Asn	Thr	Phe	Leu	Leu	Arg	Pro	Leu	His	Arg	Leu	Met	680	685	690
His	Tyr	Lys	Gln	Val	Leu	Glu	Arg	Leu	Cys	Lys	His	His	Pro	Pro	695	700	705
Ser	His	Ala	Asp	Phe	Arg	Asp	Cys	Arg	Ala	Ala	Leu	Ala	Glu	Ile	710	715	720
Thr	Glu	Met	Val	Ala	Gln	Leu	His	Gly	Thr	Met	Ile	Lys	Met	Glu	725	730	735
Asn	Phe	Gln	Lys	Leu	His	Glu	Leu	Lys	Lys	Asp	Leu	Ile	Gly	Ile	740	745	750
Asp	Asn	Leu	Val	Val	Pro	Gly	Arg	Glu	Phe	Ile	Arg	Leu	Gly	Ser	755	760	765
Leu	Ser	Lys	Leu	Ser	Gly	Lys	Gly	Leu	Gln	Gln	Arg	Met	Phe	Phe	770	775	780
Leu	Phe	Asn	Asp	Val	Leu	Leu	Tyr	Thr	Ser	Arg	Gly	Leu	Thr	Ala	785	790	795
Ser	Asn	Gln	Phe	Lys	Val	His	Gly	Gln	Leu	Pro	Leu	Tyr	Gly	Met	800	805	810
Thr	Ile	Glu	Glu	Ser	Glu	Asp	Glu	Trp	Gly	Val	Pro	His	Cys	Leu	815	820	825
Thr	Leu	Arg	Gly	Gln	Arg	Gln	Ser	Ile	Ile	Val	Ala	Ala	Ser	Ser	830	835	840
Arg	Ser	Glu	Met	Glu	Lys	Trp	Val	Glu	Asp	Ile	Gln	Met	Ala	Ile	845	850	855
Asp	Leu	Ala	Glu	Lys	Ser	Ser	Ser	Pro	Ala	Pro	Glu	Phe	Leu	Ala	860	865	870
Ser	Ser	Pro	Pro	Asp	Asn	Lys	Ser	Pro	Asp	Glu	Ala	Thr	Ala	Ala	875	880	885
Asp	Gln	Glu	Ser	Glu	Asp	Asp	Leu	Ser	Ala	Ser	Arg	Thr	Ser	Leu	890	895	900
Glu	Arg	Gln	Ala	Pro	His	Arg	Gly	Asn	Thr	Met	Val	His	Val	Cys	905	910	915
Trp	His	Arg	Asn	Thr	Ser	Val	Ser	Met	Val	Asp	Phe	Ser	Ile	Ala	920	925	930
Val	Glu	Asn	Gln	Leu	Ser	Gly	Asn	Leu	Leu	Arg	Lys	Phe	Lys	Asn	935	940	945
Ser	Asn	Gly	Trp	Gln	Lys	Leu	Trp	Val	Val	Phe	Thr	Asn	Phe	Cys	950	955	960
Leu	Phe	Phe	Tyr	Lys	Ser	His	Gln	Asp	Asn	His	Pro	Leu	Ala	Ser	965	970	975
Leu	Pro	Leu	Leu	Gly	Tyr	Ser	Leu	Thr	Ile	Pro	Ser	Glu	Ser	Glu	980	985	990
Asn	Ile	Gln	Lys	Asp	Tyr	Val	Phe	Lys	Leu	His	Phe	Lys	Ser	His	995	1000	1005
Val	Tyr	Tyr	Phe	Arg	Ala	Glu	Ser	Glu	Tyr	Thr	Phe	Glu	Arg	Trp	1010	1015	1020
Met	Glu	Val	Ile	Arg	Ser	Ala	Thr	Ser	Ser	Ala	Ser	Arg	Pro	His	1025	1030	1035
Val	Leu	Ser	His	Lys	Glu	Ser	Leu	Val	Tyr						1040	1045	

&lt;210&gt; 3

&lt;211&gt; 324

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1259001

&lt;400&gt; 3

Met	Cys	Phe	Gln	Ala	Pro	Glu	Glu	Leu	Val	Leu	His	Leu	Ala	Val
1				5					10					15
Lys	Val	Ala	Asn	Gln	Ala	Ser	Leu	Pro	Leu	Val	Asp	Phe	Ile	Ile
			20						25					30
Gln	Asn	Gly	Gly	His	Leu	Asp	Ala	Lys	Ala	Ala	Asp	Gly	Asn	Thr
			35						40					45
Ala	Leu	His	Tyr	Ala	Ala	Leu	Tyr	Asn	Gln	Pro	Asp	Cys	Leu	Lys
			50						55					60
Leu	Leu	Leu	Lys	Gly	Arg	Ala	Leu	Val	Gly	Thr	Val	Asn	Glu	Ala
			65						70					75
Gly	Glu	Thr	Ala	Leu	Asp	Ile	Ala	Arg	Lys	Lys	His	His	Lys	Glu
			80						85					90
Cys	Glu	Glu	Leu	Leu	Glu	Gln	Ala	Gln	Ala	Gly	Thr	Phe	Ala	Phe
			95						100					105
Pro	Leu	His	Val	Asp	Tyr	Ser	Trp	Val	Ile	Ser	Thr	Glu	Pro	Gly
			110						115					120
Ser	Asp	Ser	Glu	Glu	Asp	Glu	Glu	Glu	Lys	Arg	Cys	Leu	Leu	Lys
			125						130					135
Leu	Pro	Ala	Gln	Ala	His	Trp	Ala	Ser	Gly	Arg	Leu	Asp	Ile	Ser
			140						145					150
Asn	Lys	Thr	Tyr	Glu	Thr	Val	Ala	Ser	Leu	Gly	Ala	Ala	Thr	Pro
			155						160					165
Gln	Gly	Glu	Ser	Glu	Asp	Cys	Pro	Pro	Pro	Leu	Pro	Val	Lys	Asn
			170						175					180
Ser	Ser	Arg	Thr	Leu	Val	Gln	Gly	Cys	Ala	Arg	His	Ala	Ser	Gly
			185						190					195
Asp	Arg	Ser	Glu	Val	Ser	Ser	Leu	Ser	Ser	Glu	Ala	Pro	Glu	Thr
			200						205					210
Pro	Glu	Ser	Leu	Gly	Ser	Pro	Ala	Ser	Ser	Ser	Ser	Leu	Met	Ser
			215						220					225
Pro	Leu	Glu	Pro	Gly	Asp	Pro	Ser	Gln	Ala	Pro	Pro	Asn	Ser	Glu
			230						235					240
Glu	Gly	Leu	Arg	Glu	Pro	Pro	Gly	Thr	Ser	Arg	Pro	Ser	Leu	Thr
			245						250					255
Ser	Gly	Thr	Thr	Pro	Ser	Glu	Met	Tyr	Leu	Pro	Val	Arg	Phe	Ser
			260						265					270
Ser	Glu	Ser	Thr	Arg	Ser	Tyr	Arg	Arg	Gly	Ala	Arg	Ser	Pro	Glu
			275						280					285
Asp	Gly	Pro	Ser	Ala	Arg	Gln	Pro	Leu	Pro	Arg	Arg	Asn	Val	Pro
			290						295					300
Val	Gly	Ile	Thr	Glu	Gly	Asp	Gly	Ser	Arg	Thr	Gly	Ser	Leu	Pro
			305						310					315
Ala	Ser	Ser	Val	Gln	Leu	Leu	Gln	Asp						
			320											

<210> 4  
 <211> 385  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 1627027

<400> 4  
 Met Ser Val Ser Arg Thr Met Glu Asp Ser Cys Glu Leu Asp Leu  
 1 5 10 15  
 Val Tyr Val Thr Glu Arg Ile Ile Ala Val Ser Phe Pro Ser Thr  
 20 25 30  
 Ala Asn Glu Glu Asn Phe Arg Ser Asn Leu Arg Glu Val Ala Gln  
 35 40 45  
 Met Leu Lys Ser Lys His Gly Gly Asn Tyr Leu Leu Phe Asn Leu  
 50 55 60  
 Ser Glu Arg Arg Pro Asp Ile Thr Lys Leu His Ala Lys Val Leu  
 65 70 75  
 Glu Phe Gly Trp Pro Asp Leu His Thr Pro Ala Leu Glu Lys Ile  
 80 85 90  
 Cys Ser Ile Cys Lys Ala Met Asp Thr Trp Leu Asn Ala Asp Pro  
 95 100 105  
 His Asn Val Val Val Leu His Asn Lys Gly Asn Arg Gly Arg Ile  
 110 115 120  
 Gly Val Val Ile Ala Ala Tyr Met His Tyr Ser Asn Ile Ser Ala  
 125 130 135  
 Ser Ala Asp Gln Ala Leu Asp Arg Phe Ala Met Lys Arg Phe Tyr  
 140 145 150  
 Glu Asp Lys Ile Val Pro Ile Gly Gln Pro Ser Gln Arg Arg Tyr  
 155 160 165  
 Val His Tyr Phe Ser Gly Leu Leu Ser Gly Ser Ile Lys Met Asn  
 170 175 180  
 Asn Lys Pro Leu Phe Leu His His Val Ile Met His Gly Ile Pro  
 185 190 195  
 Asn Phe Glu Ser Lys Gly Gly Cys Arg Pro Phe Leu Arg Ile Tyr  
 200 205 210  
 Gln Ala Met Gln Pro Val Tyr Thr Ser Gly Ile Tyr Asn Ile Pro  
 215 220 225  
 Gly Asp Ser Gln Thr Ser Val Cys Ile Thr Ile Glu Pro Gly Leu  
 230 235 240  
 Leu Leu Lys Gly Asp Ile Leu Leu Lys Cys Tyr His Lys Lys Phe  
 245 250 255  
 Arg Ser Pro Ala Arg Asp Val Ile Phe Arg Val Gln Phe His Thr  
 260 265 270  
 Cys Ala Ile His Asp Leu Gly Val Val Phe Gly Lys Glu Asp Leu  
 275 280 285  
 Asp Asp Ala Phe Lys Asp Asp Arg Phe Pro Glu Tyr Gly Lys Val  
 290 295 300  
 Glu Phe Val Phe Ser Tyr Gly Pro Glu Lys Ile Gln Gly Met Glu  
 305 310 315  
 His Leu Glu Asn Gly Pro Ser Val Ser Val Asp Tyr Asn Thr Ser  
 320 325 330  
 Asp Pro Leu Ile Arg Trp Asp Ser Tyr Asp Asn Phe Ser Gly His

	335		340		345
Arg Asp Asp Gly	Met Glu Asp Gly Asn Lys Gln Asn Thr Asn Ser				
	350		355		360
Gln Ser Ile Gly	Ser Ile Ser Gly Gly Leu Glu Asp Gln Tyr Thr				
	365		370		375
Trp Pro Asp Thr	His Trp Pro Ser Gln Ser				
	380		385		

&lt;210&gt; 5

&lt;211&gt; 364

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1905315

&lt;400&gt; 5

Met Ser Ser Ala Pro Arg Ser Pro Thr Pro Arg Pro Arg Arg Met		
1 5 10 15		
Lys Lys Asp Glu Ser Phe Leu Gly Lys Leu Gly Gly Thr Leu Ala		
20 25 30		
Arg Lys Arg Arg Ala Arg Glu Val Ser Asp Leu Gln Glu Glu Gly		
35 40 45		
Lys Asn Ala Ile Asn Ser Pro Met Ser Pro Ala Leu Ala Asp Val		
50 55 60		
His Pro Glu Asp Thr Gln Leu Glu Glu Asn Glu Glu Arg Thr Met		
65 70 75		
Ile Asp Pro Thr Ser Lys Glu Asp Pro Lys Phe Lys Glu Leu Val		
80 85 90		
Lys Val Leu Leu Asp Trp Ile Asn Asp Val Leu Val Glu Glu Arg		
95 100 105		
Ile Ile Val Lys Gln Leu Glu Glu Asp Leu Tyr Asp Gly Gln Val		
110 115 120		
Leu Gln Lys Leu Leu Glu Lys Leu Ala Gly Cys Lys Leu Asn Val		
125 130 135		
Ala Glu Val Thr Gln Ser Glu Ile Gly Gln Lys Gln Lys Leu Gln		
140 145 150		
Thr Val Leu Glu Ala Val His Asp Leu Leu Arg Pro Arg Gly Trp		
155 160 165		
Ala Leu Arg Trp Ser Val Asp Ser Ile His Gly Lys Asn Leu Val		
170 175 180		
Ala Ile Leu His Leu Leu Val Ser Leu Ala Met His Phe Arg Ala		
185 190 195		
Pro Ile Arg Leu Pro Glu His Val Thr Val Gln Val Val Val Val		
200 205 210		
Arg Lys Arg Glu Gly Leu Leu His Ser Ser His Ile Ser Glu Glu		
215 220 225		
Leu Thr Thr Thr Thr Glu Met Met Met Gly Arg Phe Glu Arg Asp		
230 235 240		
Ala Phe Asp Thr Leu Phe Asp His Ala Pro Asp Lys Leu Ser Val		
245 250 255		
Val Lys Lys Ser Leu Ile Thr Phe Val Asn Lys His Leu Asn Lys		
260 265 270		

```

Leu Asn Leu Glu Val Thr Glu Leu Glu Thr Gln Phe Ala Asp Gly
      275                      280                      285
Val Tyr Leu Val Leu Leu Met Gly Leu Leu Glu Asp Tyr Phe Val
      290                      295                      300
Pro Leu His His Phe Tyr Leu Thr Pro Glu Ser Phe Asp Gln Lys
      305                      310                      315
Val His Asn Val Ser Phe Ala Phe Glu Leu Met Leu Asp Gly Gly
      320                      325                      330
Leu Lys Lys Pro Lys Ala Arg Pro Glu Asp Val Val Asn Leu Asp
      335                      340                      345
Leu Lys Ser Thr Leu Arg Val Leu Tyr Asn Leu Phe Thr Lys Tyr
      350                      355                      360
Lys Asn Val Glu

```

&lt;210&gt; 6

&lt;211&gt; 395

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1997789

&lt;400&gt; 6

```

Met Ser Glu Gln Ser Ile Cys Gln Ala Arg Ala Ser Val Met Val
  1          5          10          15
Tyr Asp Asp Thr Ser Lys Lys Trp Val Pro Ile Lys Pro Gly Gln
      20          25          30
Gln Gly Phe Ser Arg Ile Asn Ile Tyr His Asn Thr Ala Ser Asn
      35          40          45
Thr Phe Arg Val Val Gly Val Lys Leu Gln Asp Gln Gln Val Val
      50          55          60
Ile Asn Tyr Ser Ile Val Lys Gly Leu Lys Tyr Asn Gln Ala Thr
      65          70          75
Pro Thr Phe His Gln Trp Arg Asp Ala Arg Gln Val Tyr Gly Leu
      80          85          90
Asn Phe Ala Ser Lys Glu Glu Ala Thr Thr Phe Ser Asn Ala Met
      95          100         105
Leu Phe Ala Leu Asn Ile Met Asn Ser Gln Glu Gly Gly Pro Ser
      110         115         120
Ser Gln Arg Gln Val Gln Asn Gly Pro Ser Pro Asp Glu Met Asp
      125         130         135
Ile Gln Arg Arg Gln Val Met Glu Gln His Gln Gln Gln Arg Gln
      140         145         150
Glu Ser Leu Glu Arg Arg Thr Ser Ala Thr Gly Pro Ile Leu Pro
      155         160         165
Pro Gly His Pro Ser Ser Ala Ala Ser Ala Pro Val Ser Cys Ser
      170         175         180
Gly Pro Pro Pro Pro Pro Pro Pro Leu Val Pro Pro Pro Pro Thr
      185         190         195
Gly Ala Thr Pro Pro Pro Pro Pro Pro Leu Pro Ala Gly Gly Ala
      200         205         210
Gln Gly Ser Ser His Asp Glu Ser Ser Met Ser Gly Leu Ala Ala
      215         220         225

```





	140		145		150
Asp Met Ser Lys	Pro Trp Thr Ala Leu	Asp Ser Leu Gln Lys	Trp		
	155		160		165
Ala Ser Val Val	Arg Glu His Val Asp	Lys Leu Lys Ile Pro	Pro		
	170		175		180
Glu Glu Met Lys	Gln Met Glu Gln Lys	Leu Ile Arg Asp Phe	Gln		
	185		190		195
Glu Tyr Val Glu	Pro Gly Glu Asp Phe	Pro Ala Ser Pro Gln	Arg		
	200		205		210
Arg Asn Thr Ala	Ser Gln Glu Asp Lys	Asp Asp Ser Val Val	Leu		
	215		220		225
Pro Leu Gly Ala	Asp Thr Leu Thr His	Asn Leu Gly Ile Pro	Val		
	230		235		240
Leu Val Val Cys	Thr Lys Cys Asp Ala	Ile Ser Val Leu Glu	Lys		
	245		250		255
Glu His Asp Tyr	Arg Asp Glu His Phe	Asp Phe Ile Gln Ser	His		
	260		265		270
Ile Arg Lys Phe	Cys Leu Gln Tyr Gly	Ala Ala Leu Ile Tyr	Thr		
	275		280		285
Ser Val Lys Glu	Asn Lys Asn Ile Asp	Leu Val Tyr Lys Tyr	Ile		
	290		295		300
Val Gln Lys Leu	Tyr Gly Phe Pro Tyr	Lys Ile Pro Ala Val	Val		
	305		310		315
Val Glu Lys Asp	Ala Val Phe Ile Pro	Ala Gly Trp Asp Asn	Asp		
	320		325		330
Lys Lys Ile Gly	Ile Leu His Glu Asn	Phe Gln Thr Leu Lys	Ala		
	335		340		345
Glu Asp Asn Phe	Glu Asp Ile Ile Thr	Lys Pro Pro Val Arg	Lys		
	350		355		360
Phe Val His Glu	Lys Glu Ile Met Ala	Glu Asp Asp Gln Val	Phe		
	365		370		375
Leu Met Lys Leu	Gln Ser Leu Leu Ala	Lys Gln Pro Pro Thr	Ala		
	380		385		390
Ala Gly Arg Pro	Val Asp Ala Ser Pro	Arg Val Pro Gly Gly	Ser		
	395		400		405
Pro Arg Thr Pro	Asn Arg Ser Val Ser	Ser Asn Val Ala Ser	Val		
	410		415		420
Ser Pro Ile Pro	Ala Gly Ser Lys Lys	Ile Asp Pro Asn Met	Lys		
	425		430		435
Ala Gly Ala Thr	Ser Glu Gly Val Leu	Ala Asn Phe Phe Asn	Ser		
	440		445		450
Leu Leu Ser Lys	Lys Thr Gly Ser Pro	Gly Gly Pro Gly Val	Ser		
	455		460		465
Gly Gly Ser Pro	Ala Gly Gly Ala Gly	Gly Gly Ser Ser Gly	Leu		
	470		475		480
Pro Pro Ser Thr	Lys Lys Ser Gly Gln	Lys Pro Val Leu Asp	Val		
	485		490		495
His Ala Glu Leu	Asp Arg Ile Thr Arg	Lys Pro Val Thr Val	Ser		
	500		505		510
Pro Thr Thr Pro	Thr Ser Pro Thr Glu	Gly Glu Ala Ser			
	515		520		

<210> 8  
 <211> 348  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2363178

<400> 8  
 Met Ala Lys Leu Leu Gln Pro Pro Pro Lys Phe Leu Pro Ser Glu  
   1                  5                  10                  15  
 Trp His Ile Ala Asn Lys Asn Gln Tyr His Arg Ala Asp Ala Gln  
                   20                  25                  30  
 Arg Ser Arg Ser Glu Arg Leu Val Ala Glu Ser Gln Arg Leu Val  
                   35                  40                  45  
 Asp Glu Ile Glu Lys Thr Thr Arg Lys Ser Gln Ser Asp Val Asn  
                   50                  55                  60  
 Lys Lys Leu Glu Gln Arg Leu Glu Glu Val Gln Phe Trp Lys Lys  
                   65                  70                  75  
 Glu Leu Asp Asp Lys Leu Glu Gln Leu Val Asn Val Thr Asp Asp  
                   80                  85                  90  
 Leu Leu Ile Tyr Lys Ile Arg Leu Glu Lys Ala Leu Glu Thr Leu  
                   95                  100                 105  
 Lys Glu Pro Leu His Ile Thr Glu Thr Cys Leu Ala Tyr Arg Glu  
                  110                 115                 120  
 Lys Arg Ile Gly Ile Asp Leu Val His Asp Thr Val Glu His Glu  
                  125                 130                 135  
 Leu Ile Lys Glu Ala Glu Ile Ile Gln Gly Ile Met Ala Leu Leu  
                  140                 145                 150  
 Thr Arg Thr Leu Glu Glu Ala Ser Glu Gln Ile Arg Met Asn Arg  
                  155                 160                 165  
 Ser Ala Lys Tyr Asn Leu Glu Lys Asp Leu Lys Asp Lys Phe Val  
                  170                 175                 180  
 Ala Leu Thr Ile Asp Asp Ile Cys Phe Ser Leu Asn Asn Asn Ser  
                  185                 190                 195  
 Pro Asn Ile Arg Tyr Ser Glu Asn Ala Val Arg Ile Glu Pro Asn  
                  200                 205                 210  
 Ser Val Ser Leu Glu Asp Trp Leu Asp Phe Ser Ser Thr Asn Val  
                  215                 220                 225  
 Glu Lys Ala Asp Lys Gln Arg Asn Asn Ser Leu Met Leu Lys Ala  
                  230                 235                 240  
 Leu Val Asp Arg Ile Leu Ser Gln Thr Ala Asn Asp Leu Arg Lys  
                  245                 250                 255  
 Gln Cys Asp Val Val Asp Thr Ala Phe Lys Asn Gly Leu Lys Asp  
                  260                 265                 270  
 Thr Lys Asp Ala Arg Asp Lys Leu Ala Asp His Leu Ala Lys Ile  
                  275                 280                 285  
 Glu Gly Asn Phe Ser Pro Ser Ser Gly Arg Ala Glu Arg Ala Ala  
                  290                 295                 300  
 Ser Gln Thr Ala Cys Pro Ala Gly Gly Asp Pro Gly Gln Arg Glu  
                  305                 310                 315  
 His His Leu Tyr Arg Arg Ser Ala Val Tyr Ala Asp Glu Glu Ile  
                  320                 325                 330  
 His Pro Thr Ser Gly Trp Gly Arg Pro Trp Gly Leu Gly Trp Gly

14/35

	305		310		315
Pro Asp Ser Arg	Phe Ala Ser Leu Gln	Pro Val Ser Asn Gln	Ile		
	320		325		330
Glu Arg Cys Asn	Ser Leu Ala Phe Gly	Gly Tyr Leu Ser Glu	Val		
	335		340		345
Val Glu Glu Asn	Thr Pro Pro Lys Met	Glu Lys Glu Gly Leu	Glu		
	350		355		360
Ile Met Ile Gly	Lys Lys Lys Gly Ile	Gln Gly His Tyr Asn	Ser		
	365		370		375
Cys Tyr Leu Asp	Ser Thr Leu Phe Cys	Leu Phe Ala Phe Ser	Ser		
	380		385		390
Val Leu Asp Thr	Val Leu Leu Arg Pro	Lys Glu Lys Asn Asp	Val		
	395		400		405
Glu Tyr Tyr Ser	Glu Thr Gln Glu Leu	Leu Arg Thr Glu Ile	Val		
	410		415		420
Asn Pro Leu Arg	Ile Tyr Gly Tyr Val	Cys Ala Thr Lys Ile	Met		
	425		430		435
Lys Leu Arg Lys	Ile Leu Glu Lys Val	Glu Ala Ala Ser Gly	Phe		
	440		445		450
Thr Ser Glu Glu	Lys Asp Pro Glu Glu	Phe Leu Asn Ile Leu	Phe		
	455		460		465
His His Ile Leu	Arg Val Glu Pro Leu	Leu Lys Ile Arg Ser	Ala		
	470		475		480
Gly Gln Lys Val	Gln Asp Cys Tyr Phe	Tyr Gln Ile Phe Met	Glu		
	485		490		495
Lys Asn Glu Lys	Val Gly Val Pro Thr	Ile Gln Gln Leu Leu	Glu		
	500		505		510
Trp Ser Phe Ile	Asn Ser Asn Leu Lys	Phe Ala Glu Ala Pro	Ser		
	515		520		525
Cys Leu Ile Ile	Gln Met Pro Arg Phe	Gly Lys Asp Phe Lys	Leu		
	530		535		540
Phe Lys Lys Ile	Phe Pro Ser Leu Glu	Leu Asn Ile Thr Asp	Leu		
	545		550		555
Leu Glu Asp Thr	Pro Arg Gln Cys Arg	Ile Cys Gly Gly Leu	Ala		
	560		565		570
Met Tyr Glu Cys	Arg Glu Cys Tyr Asp	Asp Pro Asp Ile Ser	Ala		
	575		580		585
Gly Lys Ile Lys	Gln Phe Cys Lys Thr	Cys Asn Thr Gln Val	His		
	590		595		600
Leu His Pro Lys	Arg Leu Asn His Lys	Tyr Asn Pro Val Ser	Leu		
	605		610		615
Pro Lys Asp Leu	Pro Asp Trp Asp Trp	Arg His Gly Cys Ile	Pro		
	620		625		630
Cys Gln Asn Met	Glu Leu Phe Ala Val	Leu Cys Ile Glu Thr	Ser		
	635		640		645
His Tyr Val Ala	Phe Val Lys Tyr Gly	Lys Asp Asp Ser Ala	Trp		
	650		655		660
Leu Phe Phe Asp	Ser Met Ala Asp Arg	Asp Gly Gly Gln Asn	Gly		
	665		670		675
Phe Asn Ile Pro	Gln Val Thr Pro Cys	Pro Glu Val Gly Glu	Tyr		
	680		685		690
Leu Lys Met Ser	Leu Glu Asp Leu His	Ser Leu Asp Ser Arg	Arg		
	695		700		705
Ile Gln Gly Cys	Ala Arg Arg Leu Leu	Cys Asp Ala Tyr Met	Cys		
	710		715		720

Met Tyr Gln Ser Pro Thr Met Ser Leu Tyr Lys  
                             725                            730

<210> 10  
 <211> 147  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2508327

<400> 10  
 Met Pro Pro Pro Gln Lys Ile Pro Ser Val Arg Pro Phe Lys Gln  
   1                  5                  10                  15  
 Arg Lys Ser Leu Ala Ile Arg Gln Glu Glu Val Ala Gly Ile Arg  
                   20                  25                  30  
 Ala Lys Phe Pro Asn Lys Ile Pro Val Val Val Glu Arg Tyr Pro  
                   35                  40                  45  
 Arg Glu Thr Phe Leu Pro Pro Leu Asp Lys Thr Lys Phe Leu Val  
                   50                  55                  60  
 Pro Gln Glu Leu Thr Met Thr Gln Phe Leu Ser Ile Ile Arg Ser  
                   65                  70                  75  
 Arg Met Val Leu Arg Ala Thr Glu Ala Phe Tyr Leu Leu Val Asn  
                   80                  85                  90  
 Asn Lys Ser Leu Val Ser Met Ser Ala Thr Met Ala Glu Ile Tyr  
                   95                  100                 105  
 Arg Asp Tyr Lys Asp Glu Asp Gly Phe Val Tyr Met Thr Tyr Ala  
                  110                 115                 120  
 Ser Gln Glu Thr Phe Gly Cys Leu Glu Ser Ala Ala Pro Arg Asp  
                  125                 130                 135  
 Gly Ser Ser Leu Glu Asp Arg Pro Cys Asn Pro Leu  
                  140                 145

<210> 11  
 <211> 57  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2524555

<400> 11  
 Met Pro Asn Cys Arg Glu Ser Ser Phe Ser Ser Ala Thr Met Ser  
   1                  5                  10                 15  
 Asp Lys Pro Asp Met Ala Glu Ile Glu Lys Phe Asp Lys Ser Lys  
                  20                 25                 30  
 Leu Lys Lys Thr Glu Thr Gln Glu Lys Asn Pro Leu Pro Ser Lys  
                  35                 40                 45  
 Glu Thr Ile Glu Gln Glu Lys Gln Ala Gly Glu Ser  
                  50                 55

<210> 12  
 <211> 452  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2900717

<400> 12  
 Met Glu Val Thr Thr Arg Leu Thr Trp Asn Asp Glu Asn His Leu  
 1 5 10 15  
 Arg Lys Leu Leu Gly Asn Val Ser Leu Ser Leu Leu Tyr Lys Ser  
 20 25 30  
 Ser Val His Gly Gly Ser Ile Glu Asp Met Val Glu Arg Cys Ser  
 35 40 45  
 Arg Gln Gly Cys Thr Ile Thr Met Ala Tyr Ile Asp Tyr Asn Met  
 50 55 60  
 Ile Val Ala Phe Met Leu Gly Asn Tyr Ile Asn Leu Arg Glu Ser  
 65 70 75  
 Ser Thr Glu Pro Asn Asp Ser Leu Trp Phe Ser Leu Gln Lys Lys  
 80 85 90  
 Asn Asp Thr Thr Glu Ile Glu Thr Leu Leu Leu Asn Thr Ala Pro  
 95 100 105  
 Lys Ile Ile Asp Glu Gln Leu Val Cys Arg Leu Ser Lys Thr Asp  
 110 115 120  
 Ile Phe Ile Ile Cys Arg Asp Asn Lys Ile Tyr Leu Asp Lys Met  
 125 130 135  
 Ile Thr Arg Asn Leu Lys Leu Arg Phe Tyr Gly His Arg Gln Tyr  
 140 145 150  
 Leu Glu Cys Glu Val Phe Arg Val Glu Gly Ile Lys Asp Asn Leu  
 155 160 165  
 Asp Asp Ile Lys Arg Ile Ile Lys Ala Arg Glu His Arg Asn Arg  
 170 175 180  
 Leu Leu Ala Asp Ile Arg Asp Tyr Arg Pro Tyr Ala Asp Leu Val  
 185 190 195  
 Ser Glu Ile Arg Ile Leu Leu Val Gly Pro Val Gly Ser Gly Lys  
 200 205 210  
 Ser Ser Phe Phe Asn Ser Val Lys Ser Ile Phe His Gly His Val  
 215 220 225  
 Thr Gly Gln Ala Val Val Gly Ser Asp Thr Thr Ser Ile Thr Glu  
 230 235 240  
 Arg Tyr Arg Ile Tyr Ser Val Lys Asp Gly Lys Asn Gly Lys Ser  
 245 250 255  
 Leu Pro Phe Met Leu Cys Asp Thr Met Gly Leu Asp Gly Ala Glu  
 260 265 270  
 Gly Ala Gly Leu Cys Met Asp Asp Ile Pro His Ile Leu Lys Gly  
 275 280 285  
 Cys Met Pro Asp Arg Tyr Gln Phe Asn Ser Arg Lys Pro Ile Thr  
 290 295 300  
 Pro Glu His Ser Thr Phe Ile Thr Ser Pro Ser Leu Lys Asp Arg  
 305 310 315  
 Ile His Cys Val Ala Tyr Val Leu Asp Ile Asn Ser Ile Asp Asn  
 320 325 330  
 Leu Tyr Ser Lys Met Leu Ala Lys Val Lys Gln Val His Lys Glu

	335		340		345
Val Leu Asn Cys Gly Ile Ala Tyr Val		Ala Leu Leu Thr Lys Val			
	350		355		360
Asp Asp Cys Ser Glu Val Leu Gln Asp		Asn Phe Leu Asn Met Ser			
	365		370		375
Arg Ser Met Thr Ser Gln Ser Arg Val		Met Asn Val His Lys Met			
	380		385		390
Leu Gly Ile Pro Ile Ser Asn Ile Leu		Met Val Gly Asn Tyr Ala			
	395		400		405
Ser Asp Leu Glu Leu Asp Pro Met Lys		Asp Ile Leu Ile Leu Ser			
	410		415		420
Ala Leu Arg Gln Met Leu Arg Ala Ala		Asp Asp Phe Leu Glu Asp			
	425		430		435
Leu Pro Leu Glu Glu Thr Gly Ala Ile		Glu Arg Ala Leu Gln Pro			
	440		445		450
Cys Ile					

&lt;210&gt; 13

&lt;211&gt; 281

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3088904

&lt;400&gt; 13

Met Asp Ala Ile Lys Lys Lys Met Gln Met Leu Lys Glu Asn Ala		
1	5	10 15
Ile Asp Arg Ala Glu Gln Ala Glu Ala Asp Lys Lys Gln Ala Glu		
	20	25 30
Asp Arg Cys Lys Gln Leu Glu Glu Glu Gln Gln Ala Leu Gln Lys		
	35	40 45
Lys Leu Lys Gly Thr Glu Asp Glu Val Glu Lys Tyr Ser Glu Ser		
	50	55 60
Val Lys Glu Ala Gln Glu Lys Leu Glu Gln Ala Glu Lys Lys Ala		
	65	70 75
Thr Asp Ala Glu Ala Asp Val Ala Ser Leu Asn Arg Arg Ile Gln		
	80	85 90
Leu Val Glu Glu Glu Leu Asp Arg Ala Gln Glu Arg Leu Ala Thr		
	95	100 105
Ala Leu Gln Lys Leu Glu Glu Ala Glu Lys Ala Ala Asp Glu Ser		
	110	115 120
Glu Arg Gly Met Lys Val Ile Glu Asn Arg Ala Met Lys Asp Glu		
	125	130 135
Glu Lys Met Glu Leu Gln Glu Met Gln Leu Lys Glu Ala Lys His		
	140	145 150
Ile Ala Glu Asp Ser Asp Arg Lys Tyr Glu Glu Val Ala Arg Lys		
	155	160 165
Leu Val Ile Leu Glu Gly Glu Leu Glu Arg Ser Glu Glu Arg Ala		
	170	175 180
Glu Val Ala Glu Ser Arg Ala Arg Gln Leu Glu Glu Glu Leu Arg		
	185	190 195
Thr Met Asp Gln Ala Leu Lys Ser Leu Met Ala Ser Glu Glu Glu		





	50		55		60
His Val Pro Arg	Ala Val Met Ile Asp	Leu Glu Pro Thr Val Val			
	65		70		75
Asp Glu Val Arg	Ala Gly Thr Tyr Arg	Gln Leu Phe His Pro Glu			
	80		85		90
Gln Leu Ile Thr	Gly Lys Glu Asp Ala	Ala Asn Asn Tyr Ala Arg			
	95		100		105
Gly His Tyr Thr	Val Gly Lys Glu Ser	Ile Asp Leu Val Leu Asp			
	110		115		120
Arg Ile Arg Lys	Leu Thr Asp Ala Cys	Ser Gly Leu Gln Gly Phe			
	125		130		135
Leu Ile Phe His	Ser Phe Gly Gly Gly	Thr Gly Ser Gly Phe Thr			
	140		145		150
Ser Leu Leu Met	Glu Arg Leu Ser Leu	Asp Tyr Gly Lys Lys Ser			
	155		160		165
Lys Leu Glu Phe	Ser Ile Tyr Pro Ala	Pro Gln Val Ser Thr Ala			
	170		175		180
Val Val Glu Pro	Tyr Asn Ser Tyr Leu	Thr Thr His Thr Thr Leu			
	185		190		195
Glu His Ser Asp	Cys Ala Phe Met Val	Asp Asn Glu Ala Ile Tyr			
	200		205		210
Asp Ile Cys Arg	Arg Asn Leu Asp Ile	Glu Arg Pro Thr Tyr Thr			
	215		220		225
Asn Leu Asn Arg	Leu Ile Ser Gln Ile	Val Ser Ser Ile Thr Ala			
	230		235		240
Ser Leu Arg Phe	Asp Gly Ala Leu Asn	Val Asp Leu Thr Glu Phe			
	245		250		255
Gln Thr Asn Leu	Val Pro Tyr Pro Arg	Ile His Phe Pro Leu Ala			
	260		265		270
Thr Tyr Ala Pro	Val Ile Ser Ala Glu	Lys Ala Tyr His Glu Gln			
	275		280		285
Leu Ser Val Ala	Glu Ile Thr Asn Ala	Cys Phe Glu Pro Ala Asn			
	290		295		300
Gln Met Val Lys	Cys Asp Pro Arg His	Gly Lys Tyr Met Ala Cys			
	305		310		315
Cys Leu Leu Tyr	Arg Gly Asp Val Val	Pro Lys Asp Val Asn Ala			
	320		325		330
Ala Ile Ala Ala	Ile Lys Thr Lys Arg	Ser Ile Gln Phe Val Asp			
	335		340		345
Trp Cys Pro Thr	Gly Phe Lys Val Gly	Ile Asn Tyr Gln Pro Pro			
	350		355		360
Thr Val Val Pro	Gly Gly Asp Leu Ala	Lys Val Gln Arg Ala Val			
	365		370		375
Cys Met Leu Ser	Asn Thr Thr Ala Ile	Ala Glu Ala Trp Ala Arg			
	380		385		390
Leu Asp His Lys	Phe Asp Leu Met Tyr	Ala Lys Arg Ala Phe Val			
	395		400		405
His Trp Tyr Val	Gly Glu Gly Met Glu	Glu Gly Glu Phe Ser Glu			
	410		415		420
Ala Arg Glu Asp	Met Ala Ala Leu Glu	Lys Asp Tyr Glu Glu Val			
	425		430		435
Gly Ile Asp Ser	Tyr Glu Asp Glu Asp	Glu Gly Glu Glu			
	440		445		

<210> 16  
 <211> 269  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 4217506

<400> 16  
 Met Ser Asp Glu Glu Val Glu Gln Val Glu Glu Gln Tyr Glu Glu  
 1 5 10 15  
 Glu Glu Glu Ala His Glu Glu Ala Ala Glu Val His Glu Glu Val  
 20 25 30  
 His Glu Pro Glu Glu Val Gln Glu Asp Thr Ala Glu Glu Asp Ala  
 35 40 45  
 Glu Glu Glu Lys Pro Arg Pro Lys Leu Thr Ala Pro Lys Ile Pro  
 50 55 60  
 Glu Gly Glu Lys Val Asp Phe Asp Asp Ile Gln Lys Lys Arg Gln  
 65 70 75  
 Asn Lys Asp Leu Met Glu Leu Gln Ala Leu Ile Asp Ser His Phe  
 80 85 90  
 Glu Ala Arg Lys Lys Glu Glu Glu Glu Leu Val Ala Leu Lys Glu  
 95 100 105  
 Arg Ile Glu Lys Arg Arg Ala Glu Arg Ala Glu Gln Gln Arg Ile  
 110 115 120  
 Arg Ala Glu Lys Glu Arg Glu Arg Gln Asn Arg Leu Ala Glu Glu  
 125 130 135  
 Lys Ala Arg Arg Glu Glu Glu Asp Ala Lys Arg Arg Ala Glu Asp  
 140 145 150  
 Asp Leu Lys Lys Lys Lys Ala Leu Ser Ser Met Gly Ala Asn Tyr  
 155 160 165  
 Ser Ser Tyr Leu Ala Lys Ala Asp Gln Lys Arg Gly Lys Lys Gln  
 170 175 180  
 Thr Ala Arg Glu Met Lys Lys Lys Ile Leu Ala Glu Arg Arg Lys  
 185 190 195  
 Pro Leu Asn Ile Asp His Leu Gly Glu Asp Lys Leu Arg Asp Lys  
 200 205 210  
 Ala Lys Glu Leu Trp Glu Thr Leu His Gln Leu Glu Ile Asp Lys  
 215 220 225  
 Phe Glu Phe Gly Glu Lys Leu Lys Arg Gln Lys Tyr Asp Ile Thr  
 230 235 240  
 Thr Leu Arg Ser Arg Ile Asp Gln Ala Gln Lys His Ser Lys Lys  
 245 250 255  
 Ala Gly Thr Pro Ala Lys Gly Lys Val Gly Gly Arg Trp Lys  
 260 265

<210> 17  
 <211> 3620  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature

&lt;223&gt; Incyte ID No: 1285395

&lt;400&gt; 17

tgccggcggc	tgggactgaa	gagggacggg	tcccgcggcg	agcgagctcc	tgagcataag	60
ctgtggccat	gactactgaa	gtaggctctg	tgtctgaagt	gaagaaggac	tctagccagt	120
taggaacaga	tgcaaccaag	gaaaaaccta	aagaagtagc	agaaaatcag	cagaatcagt	180
cttccgatcc	agaggaggaa	aaagggtccc	agccacctcc	tgagagctgaa	agccaaagta	240
gtctacgccg	ccagaagaga	gagaaggaaa	catcgagag	caggggtatt	tctcggttca	300
taccgccatg	gcttaagaag	caaaagtcac	ataccttagt	agtggccaaa	gatggaggag	360
ataaaaaaga	gcctacccaa	gctgttgttg	aagaacaggt	cttagataaa	gaggaacccc	420
ttccagaaga	acagagacag	gctaaggggtg	atgctgaaga	aatggctcag	aagaaacaag	480
agattaaagt	tgaagtcaag	gaagaaaaac	cctcagtgag	caaggaagaa	aaaccctcag	540
tgagcaaaagt	ggagatgcag	cctactgaat	tagtaagtaa	ggagagagaa	gagaaggtaa	600
aagaaacaca	ggaagacaaa	ttagaaggag	gagcagcaaa	aagggagacc	aaggaagtgc	660
agaccaatga	gctgaaagca	gagaaggcat	ctcaaaaagt	caccaagaag	acccaaactg	720
tccagtgtaa	agtgaccctc	ttagatggca	ccgaatacag	ctgtgacctg	gagaaacatg	780
ccaagggaca	agtgttattt	gacaaagtgt	gtgaacacct	caatctcttg	gagaaagcat	840
actttggact	tttgtttcag	gaaagccctg	agcagaaaaa	ctggtttagat	cctgctaaag	900
aaataaaagag	acaactgaga	aaccttccat	ggctattcac	ttttaatgtg	aagttttatc	960
ctcctgatcc	ttctcaattg	actgaagata	tcaccagata	cttcttgtgc	cttcagctcc	1020
ggcaggacat	tgccctctggc	cgcctgccct	gctcttttgt	gactcatgct	ctcctgggat	1080
cctacaccct	gcaggctgaa	cttggtgact	atgaccacga	agaacatggc	agcatcgacc	1140
tcagtgaatt	ccagtttgcc	cctactcaga	ctaaggagct	ggaagagaag	gtggcagagc	1200
tgcacaaaac	ccacaggggc	ttatcgccag	cacaagctga	ttcccagttc	ttagaaaatg	1260
caaagagggt	ttccatgtat	ggtgttgacc	tacatcatgc	caaggactca	gaaggtgtgg	1320
acatcaagct	gggcgtgtgt	gctaattggac	ttctcattta	caaagacaga	ctgcgaatca	1380
atcgttttgc	ttggccgaaa	atcttaaaaa	tttctataaa	acgcagtaac	ttctacatta	1440
aagtcagacc	ggcagagctg	gaacagtttg	agagtaccat	tggattcaaa	ctgccaaaacc	1500
accggcgacg	gaaaagacta	tggaaagtgt	gcgtggagca	tcatactttc	tacaggcttg	1560
tttctccaga	gcagccacca	aaagccaagt	tcctgacctt	gggggtccaaa	tttcgctata	1620
gtggccgcac	ccaagcacag	acccgccagg	ccagcaccct	catagatagg	ccagcaccac	1680
actttgagcg	cacttctagt	aaacgggtct	ccaggagtct	agatggagct	ccgattgggtg	1740
tcattggacca	aagtcttatg	aaggattttc	ctggcgctgc	tggggagatt	tcagcctatg	1800
gacctggact	tgtcagcatt	gccgtggtac	aagatgggga	cggcaggagg	gaagtggagaa	1860
gccccactaa	agccccacat	ttgcagctca	ttgaaggaaa	gaaaaattcc	ttgagagtag	1920
aaggggataa	tattttatgtc	agacatagca	atttaatgtt	ggaggaaactg	gataaggccc	1980
aggaggacat	actgaaacat	caggctagca	ttagtgaact	caagcgcaat	tttatggaa	2040
ccacacctga	gcgcgcgcct	aatgaatggg	aaaaacgacg	tatcacacct	ctgtccctac	2100
agacacaagg	gagttcacat	gagactctga	atatagtggg	ggagaagaag	cgggcagagg	2160
ttgggaaaga	cgaaagagta	atcacagaag	aaatgaatgg	taaagagata	tcacctggga	2220
gtggtcctgg	ggagattcgt	aaggtggagc	ctgtgacaca	aaaagactcc	acctccctgt	2280
cttctgagag	cagcagcagc	agcagtgaga	gtgaggagga	agacgtggga	gagtaccgtc	2340
cccaccaccg	agtgaccgag	ggcaccatca	gggaggaaca	ggagtatgaa	gaagaggtgg	2400
aggaagaacc	ccgcccggca	gccaaggtag	tagagaggga	ggaagcagtg	cccgaagcca	2460
gccagtcac	acaagcaggt	gccagtgtaa	tcacagtaga	aacagtgatc	caggaaaatg	2520
taggtgcca	aaagataccc	ggagagaaga	gtgtacacga	aggcgctctt	aagcaagaca	2580
tgggagaaga	agcagaggaa	gagccacaga	aagttaacgg	agaggtgtcc	catgttgaca	2640
ttgatgtttt	gccacaaatt	atttgttgtt	cagagccacc	agtggtaaaa	acagagatgg	2700
taacaatttc	tgatgcctca	caaaggacag	aaatctccac	caaggaagtc	cccattgttc	2760
aaactgagac	caaaaccatc	acatatgagt	ctccacagat	tgatggcggg	gctgggtggg	2820
attcgggcac	gttactgacc	gcacaaacca	tcacatctga	gtccgtgtca	acaacgacaa	2880
ccacacacat	caccaagact	gtaaaagggtg	gaattttctga	aacaagaatt	gagaaacgca	2940
ttgtgatcac	aggagatgga	gatattgatc	atgaccaggc	actggctcag	gcgatcaggg	3000
aagccagaga	gcagcaccct	gacatgtcgg	tcacaagagt	ggtggtacac	aaagaaacag	3060
agttggctga	ggaaggggaa	gattaagtaa	gaaagtcatt	ttttaaacia	cactcaactt	3120

```

tgtgaacccc tgaagatttt ttgaccgttc caagtcttaa tgccacacca ctattccagc 3180
gaatztatgc tacaactggg aacaatgacc agaagcctga agaattaaaa tgccaacacc 3240
aaacctttcc ttaccagctc tgggtctatat tgctcccatg catttaatat attatattgt 3300
tttataacca cttctaaata ttctcagttc tttctttttg ttgttggtta ttaaggggtt 3360
ttggttttgt tttctgttta ctttgtgtgc aactacctgc ttttaatgac tcactttgat 3420
caaatgacag tgaacaaagc cagcccaagc tggtaagggt ctgttcactt gaacagggtc 3480
tggtgcgcag aaaggaaact ctgtgactaa tttagatagt ggctttcctt cttctggatt 3540
cttttcattg aattctcaca gtaaataatt acggagtttt caaattgcag caaatatact 3600
gtatgagaaa atattaatac 3620

```

<210> 18

<211> 4687

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 1320252

<400> 18

```

cactccagct tgggtgacag agtgagaccg ggagagggag ccgccgcagc cgccggcgct 60
gtggagatat tctctaagcc gctttcatca tgggagaaat agagcagagg ccgaccccag 120
gatcacgact gggggccccc gaaaattcgg ggatcagtag cttggaacgt ggacagaagc 180
cgcccccaac accttcagga aaactcgtgt ccatcaaaat ccagatgctg gatgacaccc 240
aggaggcatt tgaagttcca caaagagctc ctgggaaggt gctgctggat gcagtttgca 300
accacctcaa cctcgtggaa ggtgactatt ttggcctcga gtttcctgat cacaaaaaga 360
tcacggtgtg gctggatctc ctaaaaccca ttgtgaaaca gattagaagg ccaaagcacg 420
ttgttggtta gtttggtgtg aaattctttc cgctgacca cacacaactc caagaagaac 480
tcacaaggta cctgttcgcg ctgcagggtg agcaggactt ggctcaaggc aggttgacgt 540
gtaatgacac cagcgcagct ctcttgattt cacacattgt gcaatctgag attggggatt 600
ttgatgaagc cttggacaga gagcacttag caaaaaataa atacatacct cagcaagacg 660
cactagagga caaatcgtg gaatttcacc ataaccacat tggacaaaca ccagcagaat 720
cagatttcca gctcctagag attgcccgtc ggctagagat gtatggaatc cggttgaccc 780
cggccaagga cagggaaggc acgaagatca atctggccgt tgccaacacg ggaattctag 840
tgtttcaggg ttctactaag atcaatgcct tcaactgggc caaggtgcg aagctgagct 900
tcaagaggaa gcgctttttc atcaagctcc ggccagatgc caatagtgcg taccaggata 960
ccttggaatt cctgatggcc agtcgggatt tctgcaagtc cttctggaaa atctgtgttg 1020
aacatcatgc cttcttttaga ctttttgaag agcccaaacc aaagcccaag cccgtcctct 1080
ttagccgggg gtcatcattt cggttcagtg gtcggactca gaagcaggtt ctcgactatg 1140
ttaaagaagg aggacataag aaggtgcagt ttgaaaggaa gcacagcaag attcattcta 1200
tccggagcct tgcttcacag cctacagaac tgaattcgga agtgctggag cagtctcagc 1260
agagcaccag ccttacattt ggagaagggt ccgaatctcc agggggccag agctgccggc 1320
gaggaaagga accgaagggt tccgccgggg agccggggtc gcacccgagc cctgcgccga 1380
ggagaagccc cgcggttaac aagcaggcgg acggagccgc ctcggcgccc acggagggaag 1440
aggaggaggt cgttaaggat aggaccagc agagtaaacc tcagcccccg cagccaagca 1500
caggctccct gactggcagt cctcaccttt ccgagctgtc tgtgaactcg caggggggag 1560
tggccccctg caacgtgacc ttgtctccca acctgagccc cgacaccaag caggcctctc 1620
ccttgatcag cccgtgctg aatgaccagg cctgcccccg gacggacgat gaggatgagg 1680
gccggaggaa gagattccca actgataaag cgtacttcat agctaaggaa gtgtctacca 1740
ccgagcgaac atatctgaag gatctcgaag ttatcacttc gtggtttcag agcacagtga 1800
gcaaagagga cgccatgccg gaagcactga aaagtctcat attcccgaat tttgaacctt 1860
tgacaaaatt tcataaatt tttctcaagg aaattgagca acgacttgcc ctgtgggaag 1920
gccgctcaaa tgcccaaacc agagattacc aaagaatcgg cgatgtcatg ctgaagaaca 1980
ttcagggcat gaagcacctg gcggctcacc tgtggaagca cagcgaggcc ttggaggccc 2040

```

```

tggagaatgg aatcaagagc tcccggcggc tggagaactt ctgcagagac tttgagctgc 2100
agaagggtgtg ttacctaccg ctcaacacct tccctcctgcg gccactgcac cggtcatgc 2160
actacaagca ggtcctggag cggctgtgca aacaccaccc gccgagccac gccgacttca 2220
gggactgccg agccgctttg gcagagatca cggagatggg ggcacagctc caccgtacga 2280
tgatcaagat ggagaatttc cagaagctgc acgaactcaa gaaagatttg attggcattg 2340
acaatcttgt ggttccggga agggagttca tccgtctggg cagcctcagc aagctctcgg 2400
ggaaggggct ccagcagcgc atgttcttcc tgttcaacga cgtcctgcta tacacgagcc 2460
gggggctgac ggcctccaat cagtttaaaag tccacgggca gctcccgctc tatggcatga 2520
cgattgagga gagcgaagac gagtgggggg tgccccactg cctgaccctc cggggccagc 2580
ggcagtccat catcgtggcc gccagttctc ggtccgagat ggagaagtgg gttgaggaca 2640
tccagatggc cattgacctg gcggagaaga gcagcagccc cgccctgag ttcctggcca 2700
gcagccccc tgacaacaag tcccctgatg aagccaccgc ggctgaccag gattcagagg 2760
atgacctgag cgctcgcgc acatcgctgg agcgccaggc cccgcaccgc ggcaacacaa 2820
tggtgcacgt gtgctggcac cgcaacacca gcgctcccat ggtggacttc agcatcgag 2880
tggagaatca gttgtctgga aacctgctga ggaaattcaa aaacagcaac ggggtggcaga 2940
agctgtgggt ggtgttcaca aacttctgcc tgttcttcta caaatcacac caggacaatc 3000
atccccctgc cagcctgcct ctgctcggct actcgtcac catccccctc gattccgaga 3060
acatccagaa agactacgtg ttcaagctgc acttcaagtc ccacgtctac tacttcaggg 3120
cggaaagcga gtacacgttc gaaaggtgga tggaaagtga ccgcagtgc accagctctg 3180
cctcgcgacc ccacgtgttg agtcacaaag agtctcttgt gtattgatgg ccggacacac 3240
tcgtttccgc agtggctgct ttcttggaa aggtttctct tcttctgtat taatgaagcc 3300
tggtaaaatt aacacctgtc tgaaaatcaa aaacatggct tcccagcagc tctcctgtct 3360
ccacagccgc gttttttaac cccgacctct cagcgtctga atgaacagcg ctcccacctc 3420
cagtcctggc atccgctggg ggcgctgttc tttagctagt gccagtatta aaacattgtc 3480
attacgagag tgccaaatga catcttccct ccacctgcc cctgaaaaac agtacacaca 3540
catccgttca acacaagaca gggcaagtgt ttttcttctt aaaaaagtt ctttctttta 3600
ttattttcac ctattggctg ctgcatttta cgaagtggac ttcccgggtg ttgttgtttt 3660
gtttgcaata cactcagtg agccttaagc aaatgagatc attttcagat ttcatTTTTT 3720
ttttcagctc ttctactttt gtaataatag gaagttagta ggactcactt ctctgattaa 3780
taagcaattt gcagcacaca gcgttccact gcgggggttc acgtcacct gaaaacacct 3840
gttcccaacc tacttcttgg tgcaagttga ccaaactcgt ttaagtggta actctttcca 3900
accgtagcag ggttggtttc tgtaagcaa agccgagatc cagtgcata cctggactgt 3960
caccgtcctg tgagtgggtg acacaatggg aagataataa gccgtgggtg tttgctgtct 4020
gtctgtgtca caagcatgaa aaccgctgtg tcattgatca gcaccatttg tggatgttc 4080
cctgatgagc gtttagtgag cctgctggct gcagagcact atgaaatcat ggtacgtagt 4140
ccccggcacc tgtcgttatt cctatatact cctgcaactg tggtttgaaa ctgcgcattc 4200
tctagtagta tatatcgtgc ctgtcttcaa aaacatttcc ctttttatac tcattcccc 4260
caggcatggg gtagtgtcag tcggactgca cagggaacac ggtttccagt ggctttggcc 4320
cctactcggg aaacgtctgc ctgttctcga tggatggg gtggctgcca ttcccttgg 4380
tttcctaagc cttttctaac gagagtctca aacaagcgga ggcgagggcc aattcaacct 4440
cattctttcc agcgcgccgc accatagcac ctgccacct gagaaccagg aacgcacct 4500
ctctgtggag ctctgactgg tgtacctgga aacaaacagc aacttgcaaa cggacgaaga 4560
gcctgccgtg tgttaatcat ttgccttaca agatgtacca gacggtttcc agtactaaca 4620
aagggaataa aaataacctc cgccacaatc cagcatattg atgttttaag gcaaaacaaa 4680
aaaaaaa
4687

```

<210> 19

<211> 2331

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 1259001

&lt;400&gt; 19

```

gggaaatact gggttggagg caggtgcact ggggggggaa tctcagaggt gtggatgtgt 60
tttcaggcac ctgaagaact cgtcttgcac ttggctgtca aagtcgcca ccaggettcc 120
ctgcctctgg tggatttcat catccagaac ggtgggtcacc tggatgcca ggctgctgac 180
gggaacacgg ctctgcacta cgcagcactc tacaaccagc ccgactgcct caagctgctg 240
ctgaagggga gagctttggg tggcacagta aatgaagcag gcgagacagc tctggacata 300
gccaggaaga agcaccacaa ggagtgtgag gagctgctgg agcaggcca ggcggggacc 360
tttgccttcc ctctacatgt ggactactcc tgggtaattt ccacagagcc tggctctgac 420
agtgaaggag atgaggaaga gaagcgctgc ttgctgaagc tcccggcca ggctcactgg 480
gccagtggga ggctggacat cagcaacaag acctatgaga ctgtcgccag cctgggagca 540
gccacccctc agggcgagag tgaggactgt ccccgccct tggcagtcaa aaactcttct 600
cggacttttg tccaaggggtg tgcaagacat gccagtggag atcgttctga agtctccagc 660
ctgagttcag agggccctga gaccctgag agcctgggca gtccagcctc ctccctcagc 720
ctgatgagcc ccttgggaacc tggggatccc agccaagccc caccacaactc tgaagagggc 780
ctccgagagc ccccaggcac ctccagaccc agcctgacat ccgggaaccac cccttcggag 840
atgtacctcc ccgtcagatt cagctccgag agcactcgct cctatcgggc gggggcgcg 900
agccctgaag atggctccctc agccaggcag cctctgcca gaaggaacgt gccggttggc 960
atcactgaag gagatggctc aaggactggg agtctcccag caagttctgt gcaacttttg 1020
caagactagc tccttgctgg cccccacatg ccccatgcta ggccccaatg ttcagagctg 1080
ggacttgagc tcacaaaact ggggagctga gacatttgtt ctcttggtac tcaactcttc 1140
tgtcccttgt gcctctgtag ctggccttct tcctgccaca ggccatgcct ctaccaagga 1200
cacatggcct ttccctgtta gggctgatgg cggttcttct ctatctcatt acccgctagg 1260
ggcctgggag ccctgtggct ggatctgagt gctcctgagc tggcttcagc tgcagaactc 1320
tcagtccctc atcagatcga gactctattt ccccgctcag tctgggggct tcacaagggc 1380
aggagagccc tccatcactg acttccagat cagggaacct gccaaagtagg gactgtcttc 1440
tcagccagcc atttattagt ctaatatctc ttactaaat tccaactcta tgtctggacc 1500
tgtgttaggc acttcagata ccacacgagt aagacaaggg ccctgcaggg gtggtccttt 1560
ggtggaaagc tggctcttaag ggttgggctt ggggaataggc agggtcagat tccagggcat 1620
ggctctggac tcagctgggt tatacctata tgaccattac agttgtctac agatcacatc 1680
cattctggct ggtcaacatg catgctgtac tggctgttaa ataaaaatat tctgaatgtc 1740
actccttttg agggacagca cagccttccc taggcattct cctatatctc ccagccaaat 1800
tgtagagtca gatgcacca catttgcttg tgccttgat ttagcaggaa ggaaaggaat 1860
agtcgggggt gatggatgcc cacttctctt ctctttctct tgggtcaactc aggagccttt 1920
tagtctgagg gaatggagag gcaaagaaag aaggagagat aatagaattg ggagggcaga 1980
gacttaaggg ttctgcttcc cagccctaga aattctatca ttgctcagcc ccaatgagaa 2040
agcagataca cctaagccat catcaaccac taacatctca acttgccagt tgctgggtgc 2100
tgggccctgg caggaatggg ccaagccaag caggggagac tagagagcac caatggccaa 2160
cacagctgcc tggctgggga ggctgtgctg tttccctgg agacctgact ggtctgtggc 2220
tcccacagga acagggttgt cttttgagcc ccagtgctc ggtttcattc atctcagact 2280
tgttatttca ctcatctcta ataaaggatt ggggggtcaa aaaaaaaaaa a 2331

```

&lt;210&gt; 20

&lt;211&gt; 1918

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1627027

&lt;400&gt; 20

```

cagtggccct gggaccagct ctgctccttg caccocgctc cctgcctgga cacaggctca 60
ctcgctgcct tcttctgggg gaaaccagct tcttgccagc cacagctgct gcctccgcca 120
ctggccaccg cccctgtcct gggagtccct tggcccagac acccacctga cttagtggct 180

```

```

cctctgcagg aaaggtggct gccccctgcg ttcctccatc caaccatgag ctggtgcccc 240
tcaccactga gaatgcacca aagaatgtag tggacaaggg agaaggagcc tccccgggtg 300
gaaacacacg gaaaagcctc gaggacaacg gctccaccag ggtcaccocg agtgtccagc 360
cccacctcca gcccatcaga aacatgagtg tgagccggac catggaggac agctgtgagc 420
tggacactgg gtacgtcaca gagaggatca tcgctgtctc cttccccagc acagccaatg 480
aggagaactt ccggagcaac ctccgtgagg tggcgcagat gctcaagtc aaacatggag 540
gcaactacct gctgttcaac ctctctgagc ggagacctga catcacgaag ctccatgcca 600
aggtactgga atttggctgg cccgacctcc acacccagc cctggagaag atctgcagca 660
tctgtaaggc catggacaca tggctcaatg cagacctca caatgtcgtt gttctacaca 720
acaagggaaa ccgaggcagg ataggagttg tcatcgcggc ttacatgcac tacagcaaca 780
tttctgccag tgcggaccag gctctggacc ggtttgcaat gaagcggttc tatgaggata 840
agattgtgcc cattggccaag ccatcccaaa gaaggtagct gcttacttc agtggcctgc 900
tctccggctc catcaaatg aacaacaagc ccttgtttct gcaccacgtg atcatgcacg 960
gcatccccaa ctttgagtct aaaggaggat gtcggccatt tctccgcac taccaggcca 1020
tgcaacctgt gtacacatct ggcattctaca acatcccagg agacagccag actagcgtct 1080
gcatcaccat cgagccagga ctgctcttga agggagacat cttgctgaag tgctaccaca 1140
agaagttccg aagcccagcc cgagacgtca tcttcggtg gcagttccac acctgtgcca 1200
tccatgacct gggggttgtc tttgggaagg aggaccttga tgatgctttc aaagatgatc 1260
gatttccaga gtatggcaaa gtggagtttg tattttctta tgggcccagag aaaattcaag 1320
gcatggagca cctggagaac gggccgagcg tgtctgtgga ctataacacc tccgaccccc 1380
tcacccgtg ggactcctac gacaacttca gtgggcatcg agatgacggc atggaggatg 1440
ggaacaaaca gaataccaat agccagtcaa ttggaagcat ctcagggtggc ctggaggatc 1500
aatataacct gccagacaca cactggcctt cccaaagtta ggccctccaa aatgtcaagt 1560
tgccaagcca gaatccaaag ctttgcaactg agaacttcat gttggggctc ctgtcccttc 1620
ctggagccct gcctgggtga cccccagtc cttctattca ggtcctgtgt tcacctctgg 1680
gaagtgcagg accctgtggg gttttttctt tttccttttc caagccaagg aagtgttctg 1740
gctcctggag tctgccctgt ggaggggaaa gggggagaga gagggaaaaa aagatttcca 1800
ccttattgtt cagcagctgg aaaattgttg tcatctgttc tgccttccat ttgctgagaa 1860
tggaaaaaaa atgtggaaaa tttggggaaa taaatctcaa taccctcaaa aaaaaaaa 1918

```

<210> 21

<211> 1386

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 1905315

<400> 21

```

cgggcggtc cacacgcgct gcgcccgcg cgggccccac gcgcggccca tgtcctccgc 60
gccgcgctcg cccaccccg gcgcccgcag gatgaagaag gacgagtcgt tcctgggcaa 120
gctggcgcg accctggcca ggaagcggag ggcgcgcgag gtgagtgacc tgcaggaaga 180
aggcaagaat gccatcaact caccgatgtc cccgcgcctg gcgatgttc accctgaaga 240
caccagctc gaggagaacg aggagcgcac gatgattgac ccacttcca aggaagacc 300
caagtccaag gaactggtca aggtcctcct cgactggatt aatgacgtgc tggaggagga 360
gaggatcatt gtgaagcagc tggaggaaga cctgtatgac ggccagggtc tgcagaagct 420
cttggaaaaa ctggcagggg gcaagctgaa tttggctgag gtgacacagt ccgaaatagg 480
gcagaaacag aagctgcaga cgggtctgga agcagtagat gacctgctgc ggccccgagg 540
ctggcgctc cgggtggagcg tggactcaat tcacgggaag aacctggtgg ccatcctcca 600
cctgctggtc tctctggcca tgcacttcag ggcccccatc cgccttctct agcatgtaac 660
ggtgcagggt gtggtcgtgc ggaaacggga aggcctgctg cattccagcc acatctcgga 720
ggagctgacc acaactacag agatgatgat gggccgggtc gagcgggatg ccttcgacac 780
gctgttcgac cacgccccgg ataagctcag cgtgggtgaag aagtctctca tcacttttgt 840

```



```

gaacaagcac ctgaacaagc tgaatttggg ggtgacggaa ctggagaccc agtttgcaga 900
tggcgtgtac ctggttctgc tcatgggctt tctggaagac tactttgttc ctctccacca 960
cttctacctg actccggaaa gcttcgatca gaaggctccac aatgtgtcct tcgcctttga 1020
gctgatgctg gacggaggcc tcaagaaacc caaggctcgt cctgaagacg tggttaactt 1080
ggacctcaaa tccaccctga gggttcttta caacctgttc accaagtaca agaacgtgga 1140
gtgacggggg agctgtggat ggtggcagga gtgtccagc aagaaaggcg gcatccgtct 1200
gtgccctgtg cctttccagg gagccaggcg ccatgggctt ctggtccaag ctgtgttgac 1260
tgtcatcccc accccacccc tacctcacgc ctgccccacc cctgcctct tttggttgtt 1320
gttcttaatc tcctctccat gtagttccca gtgggcaaga gcctttgaaa atgcaggatt 1380
ctaaaa                                           1386

```

<210> 22

<211> 1745

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 1997789

<400> 22

```

ccgggtcgcc cctggccccg cgcgcccgct cccggcagcg acaatgagtg aacagagtat 60
ctgccaagcc cgggcttccg tgatggtcta cgatgacacc agtaagaaat gggtagcaat 120
caaacctggc cagcagggat tcagccggat caacatctac cacaacactg ccagcaacac 180
cttcagagtc gttggagtca agttgcagga tcagcagggt gtgatcaatt attcaatcgt 240
gaaagggctg aagtacaatc agggcacgcc aaccttcac cagtggcgag atgcccgcca 300
ggtctacggc ttaaactttg caagtaaaga agaggcaacc acgttctcca atgcaatgct 360
gtttgccctg aacatcatga attcccaaga aggaggcccc tccagccagc gtcagggtgca 420
gaatggcccc tctcctgatg agatggacat ccagagaaga caagtgatgg agcagcacca 480
gcagcagcgt caggaatctc tagaaagaag aacctcggcc acagggccca tcctcccacc 540
aggacatcct tcactctgag ccagcgcccc cgtctcatgt agtgggcctc caccgcccc 600
cccacctcta gtcccacct caccactgg ggtacccca cctccccac cccactgcc 660
agccggagga gccccagggg ccagccacga cgagagctcc atgtcaggac tggccgctgc 720
catagctggg gccaaagctg gaagagtcca acggccagaa gacgcactcg gaggctccag 780
tcccagtgga acctcaaagt ccgatgcaa ccgggcaagc agcgggggtg gcggaggagg 840
cctcatggag gaaatgaaca aactgctggc caagaggaga aaagcagcct ccagtcaga 900
caagccagcc gagaagaagg aagatgaaag ccaaattgaa gatcctagta cctccccctc 960
tccggggacc cgagcagcca gccagccacc taactcctca gaggctggcc ggaagccctg 1020
ggagcggagc aactcgggtg agaagcctgt gtcctcgatt ctgtccagga tgaagcctgc 1080
tgggagcgtg aatgacatgg ccctggatgc ctctgacttg gaccggatga agcaggagat 1140
cctagaggag gtggtgagag agctccaca ggtgaaggag gagatcatcg acgccatcag 1200
gcaggagctg agtgggatca gcaccacgta aggggccggc ctgctgcgc tgattcgctg 1260
agcccatccg gcgacagagg acagccagaa gccagccag cccagactc cagtgcacca 1320
gagcacgcac aggagcctgg gcgcgctgct gtgaaacgtc ctgacctgtg atcacacatg 1380
acagtgagga aaccaagtgc aactcctggg tttttttaga ttctgcctga cacggaacac 1440
caggtctgct cgtctttttt gtgttttata tttgcttatt taaggtacat ttctttgggt 1500
ttctagagac gcccctaagt cacctgcttc attagcgggt ttccagggtt tctcccaggt 1560
gacgctgtta gcgcctcagc tggcggtgac agccggccca gcgtggcgcc accacacacc 1620
gcagagctgt ccaggcacag ctccgtcccc agcgtcatg gtgttgaaac tgtctgtcat 1680
gcaccacggg gtctgtgtcc acacagtaat aaacggttta ctgtccgcaa aaaaaaaaaa 1740
aaagg                                           1745

```

<210> 23  
 <211> 2451  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2303465

<400> 23  
 cgccctcagcc gcctcgcaca tttagtcttg ccgggagtggt tgtgattccc gaccaagatg 60  
 gcggccgtgg ggcgagtcgg ctcccttcgggt tcttctccgc cgggattatc ctcgacttac 120  
 actggcgggcc ccttgggcaa cgagatagcg tcgggcaacg gtggcgccgc ggcaggcgac 180  
 gacgaggacg ggcagaacct ttggtcctgc atcctcagcg aggtctccac ccgctcgcgc 240  
 tccaagctcc ctgcggggaa gaacgtgcta ctgctgggtg aagatggagc tggaaaaaca 300  
 agcttaataa gaaaaattca gggaaatagag gagtataaga aaggaagagg attggaatat 360  
 ttgtacttaa atgtgcatga tgaagacagg gatgatcaaa caagatgtaa tgtttggatc 420  
 ttagatggag acctatatca caaaggcctc cttaaatttt cactggatgc cgtatctctg 480  
 aaggatactc tagttatgct ggttggtgac atgtcaaagc cttggactgc tttggattct 540  
 ttacagaaat gggcaagtgt tgttagagaa catggtgaca aactgaaaat ccctcctgaa 600  
 gaaatgaaac aaatggaaca aaagttgatt agagacttcc aagaatatgt agagccagga 660  
 gaagacttcc cggtcttctc ccagagaaga aatactgcgt cacaagaaga caaagatgac 720  
 agtgtagttt tacctctggg tgcggataca cttacacata acttgggcat tccagtacta 780  
 gtagtttgca caaagtgtga tgccattagt gtattggaga aagaacatga ctacagagat 840  
 gaacattttg attttattca gtcacatatc cggaagtttt gtttacagta tgggtgcagca 900  
 cttattttaca cttcagtaaa agaaaacaaa aatatagact tagtatataa atacatcggt 960  
 cagaaactat atggatttcc ctataagatt cctgctgttg ttgtggaaaa ggatgcagta 1020  
 tttattccag cagggtggga taatgataag aaaataggaa tattacatga aaattttcaa 1080  
 acattaaaag cagaagataa ttttgaagac atcataacta aaccacctgt tcgaaaagttt 1140  
 gtacatgaga aggaaattat ggcagaagat gatcagggtgt ttcttatgaa gctacagtcc 1200  
 ctttttagcaa agcaaccacc aactgcagct ggaaggcctg tggatgcctc accaagagtc 1260  
 ccaggaggct cccacgaac accaaataga tctgtatcat ctaatgttgc cagcgtgtca 1320  
 cccattcctg ctgggtcaaa aaaaattgat ccaaactga aagctggagc tacaagtga 1380  
 ggcgttcttg caaatttctt caacagtttg ttgagtaaaa agactggctc tccaggaggc 1440  
 cctggtgtga gtggtggtag ccctgcaggt ggggctggag gtggaagcag tggtttacca 1500  
 ccatccacca aaaagtcagg ccagaagcct gtcttagatg ttcattgcaga actagacaga 1560  
 attacacgaa aaccagttac agtttctccc acaacaccta catctcctac ggaaggagaa 1620  
 gcttcttgaa gataccaaat aaagccattt attctgtttt ctgggataat gtaaactatgc 1680  
 ctctgccttt tcttcaaaa gtggaattag aaagctggag tgcttcttca gatggactaa 1740  
 atttatgtcg tgtgtgtgtg tgtgtgtgtg tgtggttacc catttttttag aaggagccgt 1800  
 acagaagaaa aattattcta cattatgtag gattgctgtt tgcattgcca ttttgcataa 1860  
 gaaagtaatt tttgattttg aaaatctcaa aactttttaga tctgaaatac agccatgtga 1920  
 tcgatcatat tctaaaggct atttaaaaca tgtaaaagga tttggggaac ggcagaaaac 1980  
 atgcagtttg ggcatttgac tgacttgga gtctaagctt attttagtta taactattaa 2040  
 aatcattttt aaaaatttgt tagttttgct gacagagaaa aatcgtcagt tgtcagatgt 2100  
 tgcacaccac aataaatgta ccaccacaca cggaatatgc taagaaaact atcagatagc 2160  
 gtttgatata ctagtcattg tctcaatcac tgatcctgta agttgtcatc aaaatatgat 2220  
 ttagaaatat tggccaaggt gttgctttta ctgaggagaa aagaaagcac actgcctaaa 2280  
 tgtgtaaaag aaaaatgcag aggttattaa aatgtaaaaga agtaacaatc tttggatttg 2340  
 tctatacata tatatatata tatatggctt tgcttaata taccctttt tttgtttgtg 2400  
 actttcaact gtaatcagtt aataaagtat ttattctctg caaaaaaaaa a 2451

<210> 24  
 <211> 1162

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2363178

&lt;400&gt; 24

```

ggagggctccc atccccgagct tgcgtagtgg ttctgtgtgg cgctgaggaa tttggaaacc 60
tcaaattggct aaactattac aacctccgcc caagttcctg ccctcagagt ggcacattgc 120
taacaagaac cagtagcaca gagcagacgc tcaaagggtcc cgatcagaac gcctgggtcgc 180
agaaagccag aggcttgtgg atgaaattga aaagaccaca agaaaatctc aaagcgatgt 240
gaacaagaaa ctagaacaga gactcgagga agtccagttc tggaagaagg agttagatga 300
caaacttgag cagcttgtga atgtaactga tgatctactc atatataaga tcagattgga 360
aaaagccctg gagaccttga aagagccctt gcacatcact gagacatgcc tggcatacag 420
ggagaagcgc attggcattg acctggtgca cgacacagtg gagcatgagc tgataaagga 480
ggctgagatc atccagggca ttatggctct gctgaccctg acctggagg aggcttccga 540
gcagattcgg atgaaccgct ctgccaagta caatcttgag aaggatttga aggacaagtt 600
tgtggccctg accatagatg atatctgctt ctgctcaac aacaactcac caaacatcag 660
atattctgag aacgccgtga ggattgagcc aaactccgtg agtctggaag actggttga 720
cttctccagc accaatgtgg agaaggctga caagcagcgg aacaactccc tgatgctgaa 780
agccctggtg gatcgaatcc tgtcccagac agccaatgat ctgcgcaagc agtgtgatgt 840
ggtggacacg gcattcaaga atgggctgaa ggatacaaag gatgccaggg acaagctggc 900
tgatcatctg gccaaagatt aaggaaactt tagcccaagc tcaggcagag ctgaaagggc 960
tgcacgcag acagcttgcc ctgcaggagg agatccaggt caaagagaac accatttata 1020
tcgacgaagt gctgtgtatg cagatgagga aatccatccc acttcgggat ggggaagacc 1080
atggggtctg ggctgggggc ctccgccctg atgctgtctg ctaatagtag ggctagttcc 1140
aattctcatt aaaccgggat ga 1162

```

&lt;210&gt; 25

&lt;211&gt; 4527

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2363327

&lt;400&gt; 25

```

agccccgggac gcgggctggg gagccggggc gagggggcgac gccccgccgc ccgagtttcc 60
ccctttctag ggtgaggatg gttctacaca gccacccgga gttccttagt tgaaagggtgc 120
gcctgctgt gacagcatgg acaccacgtt gctgaaaaca tgctttggga ctgccactga 180
atttatcttt tgcggtttta tgacaaagtt attagtagtt tccctttttt gaattagtat 240
tttgaagtta atatcacaat gagtttccag gcttatggag ccaagaaaaa gtcacttcac 300
cctactggga agagcggatt ttttacttgc ttcttcaaga atgcagcgtt acagacaaac 360
aaacacaaaa gtccttataa gtaccgaagg aagtatagga cagtatatcc aagatcggtc 420
tgtggggcat tcaaggattc cttctgcaaa aggcaagaaa aatcagattg gattaaaaat 480
tctagagcaa cctcatgcag ttctctttgt tgatgaaaag gatgtttag agataaatga 540
aaagttcaca gagttacttt tggcaattac caattgtgag gagaggttca gcctgtttaa 600
aaacagaaac agactaagta aaggcctcca aatagacgtg ggctgtcctg tgaaagtaca 660
gctgagatct ggggaagaaa aatttctctg agttgtacgc ttcagaggac ccctgttagc 720
agagaggaca gtctccgga tattcttttg agttgaattg ctggaagaag gtcgtggtca 780
aggtttcaat gacggggtgt accaagggaa acagcttttt cagtgtgatg aagattgtgg 840
cgtgtttgtt gcattggaca agctagaact catagaagat gatgacactg cattggaaa 900

```

```

tgattacgca ggtcctgggg acacaatgca ggtcgaactt cctccttttg aaataaaactc 960
cagagtttct ttgaagggtg gagaaacaat agaactctgga acagttatat tctgtgatgt 1020
tttgccagga aaagaaagct taggatattt tgttggtgtg gacatggata accctatttg 1080
caactgggat ggaagatttg atggagtgc gctttgtagt tttgcgtgtg ttgaaagtac 1140
aattctattg cacatcaatg atatcatccc agagagtgtg acgcaggaaa ggaggcctcc 1200
caaacttgcc tttatgtcaa gaggtgttgg ggacaaaagg tcatccagtc ataataaacc 1260
aaaggctaca ggatctacct cagaccctgg aaatagaaac agatctgaat ttttttatac 1320
cttaaatggg tcttctgttg actcacaacc acaatccaaa tcaaaaaata catggtacat 1380
tgatgaagtt gcagaagacc ctgcaaaatc tcttacagag atatctacag actttgaccg 1440
ttcttcacca ccaactccagc ctctcctgt gaactcactg accaccgaga acagattcca 1500
ctctttacca ttcagtctca ccaagatgcc caataccaat ggaagtattg gccacagtcc 1560
actttctctg tcagcccagt ctgtaatgga agagctaaac actgcacccg tccaagagag 1620
tccacccttg gccatgcctc ctgggaactc acatggtcta gaagtgggct cattggctga 1680
agttaaggag aaccctcctt tctatggggt aatccgttgg atcggtcagc caccaggact 1740
gaatgaagtg ctcgctggac tggaaactgga agatgagtgt gcaggctgta cggatggaac 1800
cttcagaggc actcgggtatt tcacctgtgc cctgaagaag gcgctgtttg tgaaactgaa 1860
gagctgcagg cctgactcta gggttgcatc attgcagccg gtttccaatc agattgagcg 1920
ctgtaactct ttagcatttg gaggtactt aagtgaagta gtagaagaaa atactccacc 1980
aaaaatggaa aaagaaggct tggagataat gattgggaag aagaaaggca tccagggtca 2040
ttacaattct tgttacttag actcaacctt attctgctta tttgctttta gttctgttct 2100
ggacactgtg ttacttagac ccaaagaaaa gaacgatgta gaattattata gtgaaacca 2160
agagctactg aggacagaaa ttgttaatcc tctgagaata tatggatatg tgtgtgccac 2220
aaaaattatg aaactgagga aaatacttga aaagggtggag gctgcacag gatttacctc 2280
tgaagaaaaa gatcctgagg aattcttgaa tattctgttt catcatattt taagggtaga 2340
acctttgcta aaaataagat cagcaggtca aaagggtacaa gattgttact tctatcaaat 2400
ttttatggaa aaaatgaga aagttggcgt tcccacaatt cagcagttgt tagaatggc 2460
ttttatcaac agtaacctga aatttgaga ggcaccatca tgtctgatta ttcagatgcc 2520
tcgatttgga aaagacttta aactatttaa aaaaattttt ccttctctgg aattaaatat 2580
aacagattta cttgaagaca ctcccagaca gtgccggata tgtggagggc ttgcaatgta 2640
tgagtgtaga gaatgctacg acgatccgga catctcagct ggaaaaatca agcagttttg 2700
taaaacctgc aacactcaag tccaccttca tccgaagagg ctgaatcata aatataacct 2760
agtgtcactt ccaaagact taccgactg ggactggaga caggctgca tcccttgcca 2820
gaatatggag ttatttgctg ttctctgcat agaaacaagc cactatgttg cttttgtgaa 2880
gtatgggaag gacgattctg cctggctctt ctttgacagc atggccgatc gggatgggtg 2940
tcagaatggc ttcaacattc ctcaagtcac cccatgcccc gaagtaggag agtacttgaa 3000
gatgtctctg gaagacctgc attccttgga ctccaggaga atccaaggct gtgcacgaag 3060
actgctttgt gatgcatata tgtgcatgta ccagagtcca acaatgagtt tgtacaaata 3120
actggggtca tcgggaaagg caaagaaact gaagcagag tcctaacgtt gcatcttatt 3180
cgagctggca gttctgttca cgtccattgc aggcaatgga tgtctttgtg gtgatgatcc 3240
ttcagaaaag gatgcctctg tttaaaaaca aattgctttt gtgtccctga agtatttaat 3300
aagaagcatt ttgactcta gaaagtatgt ttgtgttggg tttttaagaa gtctaaatga 3360
agttattaat acctgaagct ttaagttaag tgcattgatc atatgatatt tttggaagca 3420
tacaatttta attgtggaag tttaaagcct cttttagtcc attgagaatg taaataaatg 3480
tgtcttcttt atggaccaag gatatgaaat catttttctt ttgtagctaa cggttgcctt 3540
gaggaagaaa taatttggtt ttattaagag tctactctca atccagttat tagagatgta 3600
ctgagtttga tttgttaatc ctttctatat actgctgatc ttgcatgtct acaatctgct 3660
cagtttttct gtgtttctgc aatagtgggtc agaaaaatac ttaaatccc ttaatgggtg 3720
tgttttctat ttgttctggg tttgagataa atgagtgtt tttagcagtt gaagctcttc attcatagta 3780
ttgaagtgat tttcctggag gattagggtg ttttaactat taatatatta tgggaatggg 3900
gttactgtca gctaacaggt ttttaagggc ttttaactat taatatatta tgggaatggg 3960
caaagtaaat tgaatgaaaga attggagtga taatagtcct ttacaaacat acagtccata 3960
agaaaatgaa tttggcatat agaattatta caatttctct ggagagatgg atatttaaac 4020
ctctattatt ttagacaaga ctgtctagaa cttaagtttg atctgtcagc cagtactccc 4080
attaaattca gtgtagtttc acttgataga atcagatatg ttatcgaaat gttagcagca 4140
gcttcatcct ccttctgatt aaagtaagta gaaatgggat gttttgttta ataacagcca 4200

```

```

tagtgtgtgt ttagaccaca gcggatgttg tagaccagga ccatagatga tacatgtcag 4260
tgctgtggaa tgtgcattct ctgagtgttg ttttgtggta tcattgtctt tcctgaatga 4320
ctttctaact gtgcagaaag gcagaaaagt catcatatgt atatgtcata tgactttata 4380
aaatatattaa tgtgacaaaa agtggaaaga atctttacaa accctgcaat tactttttta 4440
aaggcacttt tactcttttg ttttatcatt ccattttgct aatatttact agctttataa 4500
attacagtaa ggtacaaaaa aaaaaaa 4527

```

<210> 26  
 <211> 870  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2508327

```

<400> 26
ggaggaatga gttaggttcc cgggtgctgg acagtttttt tttctttttt aaaacagaca 60
cagctactga gtgcaatgcc gcctccacag aaaatcccaa gcgtcagacc cttcaagcag 120
aggaaaagct tggcaatcag acaagaggaa gttgctggaa tccgggcaaa gttcccaaac 180
aaaatcccgg tggtagtgga gcgctacccc agggagacgt tcctgcccc gctggacaaa 240
accaagttcc tggccccga ggagctgacc atgaccagc tcctcagcat catccggagc 300
cgcattggtc tgagagccac ggaagccttt tacttgctgg tgaacaacaa gagcctggc 360
agcatgagcg caaccatggc agagatctac agagactaca aggatgagga tggcttcgtg 420
tacatgacct acgcctccca ggagacattt ggctgcctgg agtcagcagc cccagggat 480
gggagcagcc ttgaggacag accctgcaat cctctctagc ccatgtcggg aaggatgtgt 540
gctctgacag acgtgtcaga tgctggcaga agggattggg tttctcctgt gtatacagtg 600
gaggagtctg agaagcaggg atgcctgggg tgatcagctc caaccagtgg cagcagagtg 660
gtggctcttt cctagtttag tttttttt cgtttttttg tgccttcta agaaaaatgt 720
gggcccggcc tgaaagggtat ataattatca tttttttttt tttgaaacag ggtcttgtct 780
tactgccttg gttggagcac agtgttgctt tggattgtac ctactgtact tccatcctgt 840
gcttttaatg tgactcgtg tacttgaaga 870

```

<210> 27  
 <211> 729  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2524555

```

<400> 27
gtggccactg cgcagaccag acttcgctcg tactcgtgcg cctcgtctcg gtgagcccca 60
gggcccctgc ctcccttctc ctgccgtcct gcctccgtcc ccgccctttc atcatccgcg 120
tccctgtgaa ggcattccct aaatccgagc ccgagtgggt ctccccggga aggctacttt 180
ggggagctgg ggggatgcca aacaccctag atactggata atgggggtgg gaaatcgatg 240
atttaagaac aaaaccgaaa aactggcggt ttgtcgtgcc gctcggaggg gacattaaaa 300
aatttcttag tgtttgcccg caaagggtat gtgcgttgcc ttggaggctg agatatgggg 360
gaatagacaa gtcctttgtt ctgaggttca tcttccgagc cccgagcctc ctccagcct 420
cggacggctg cgcgggctgc atctgtgcag cctggcggtg gcggggctgt gctatgacat 480
ctttacagtc cttcttgacg agacatgtgt gccagggatg ccgaattgcc gggagagcag 540
cttttctctc gcaaccatgt ctgacaaacc cgatatgggt gagatcgaga aattcgataa 600

```

```

gtcgaaactg aagaagacag agacgcaaga gaaaaatcca ctgccttcca aagaaacgat 660
tgaacaggag aagcaagcag gcgaatcgta atgaggcggtg cgccgccaat atgcactgta 720
cattccaca                                     729

```

&lt;210&gt; 28

&lt;211&gt; 2062

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2900717

&lt;400&gt; 28

```

gagtctctga agccacagat ctcttaagaa ctttctgtct ccaaaccgtg gctgctcgat 60
aatcagaca gaacagttaa tcctcaattt aagcctgatc taacccttag aaacagatat 120
agaacaatgg aagtgacaac aagattgaca tggaaatgatg aaaatcatct gcgcaagctg 180
cttggaatg tttctttgag tcttctctat aagtctagtg ttcattggagg tagcattgaa 240
gatatgggtg aaagatgcag ccgtcagggg tgtactataa caatggctta cattgattac 300
aatatgattg tagcctttat gcttggaaat tatattaatt tacgtgaaag ttctacagag 360
ccaaatgatt ccctatgggt ttcacttcaa aagaaaaatg acaccactga aatagaaact 420
ttactcttaa atacagcacc aaaaattatt gatgagcaac tgggtgtgtcg tttatcgaaa 480
acggatattt tcattatatg tcgagataat aaaatttatc tagataaaat gataacaaga 540
aacttgaaac taaggtttta tggccaccgt cagtatttgg aatgtgaagt ttttcgagtt 600
gaaggaatta aggataaacct agacgacata aagaggataa ttaaagccag agagcacaga 660
aataggcttc tagcagacat cagagactat aggccctatg cagacttggg ttcagaaatt 720
cgtattcttt tgggtgggtcc agttgggtct ggaaagtcca gttttttcaa ttcagtcaag 780
tctatttttc atggccatgt gactggccaa gccgtagtgg ggtctgatac caccagcata 840
accgagcggg ataggatata ttctgttaaa gatggaaaaa atggaaaatc tctgccattt 900
atgttgtgtg acactatggg gctagatggg gcagaaggag caggactgtg catggatgac 960
attccccaca tcttaaaagg ttgtatgcca gacagatatc agtttaattc ccgtaaacca 1020
attacacctg agcattctac ttttatcacc tctccatctc tgaaggacag gattcactgt 1080
gtggcttatg tcttagacat caactctatt gacaatctct actctaaaat gttggcaaaa 1140
gtgaagcaag ttcacaaaga agtattaaac tgtgggtatag catatgtggc cttgcttact 1200
aaagtggatg attgcagtga gggtcttcaa gacaactttt taaacatgag tagatctatg 1260
acttctcaaa gccgggtcat gaatgtccat aaaatgctag gcattcctat ttccaatatt 1320
ttgatgggtg gaaattatgc ttcagatttg gaactggacc ccatgaagga tattctcatc 1380
ctctctgcac tgaggcagat gctgcgggct gcagatgatt ttttagaaga tttgcctctt 1440
gaggaaactg gtgcaattga gagagcggtt cagccctgca tttgagataa gttgccttga 1500
ttctgacatt tggcccagcc tgtactgggt tgcgcgaatg agagtcaatc tctattgaca 1560
gcctgcttca gattttgctt ttgttcgttt tgccttctgt ccttggaaaca gtcatatctc 1620
aagttcaaaag gccaaaacct gagaagcggg gggctaagat aggtcctact gcaaaccacc 1680
cctccatatt tccgtacat ttacaattca gtttctgtga catcttttta aaccactgga 1740
ggaaaaatga gatattctct gagatattct ctaatttatt cttctataac actctatata 1800
gagctatgtg agtactaatc acattgaata atagttataa aattattgta tagacatctg 1860
cttcttaaac agattgtgag ttctttgaga aacagcgtgg attttactta tctgtgtatt 1920
cacagagctt agcacagtgc ctggtaatga gcaagcatac ttgccattac ttttccttcc 1980
cactctctcc aacatcacat tcactttaaa tttttctgta tatagaaagg aaaactagcc 2040
tgggcaacat gatgaaaccc ct                                     2062

```

&lt;210&gt; 29

&lt;211&gt; 1020

&lt;212&gt; DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 3088904

<400> 29

```

ctcgccctgcc accggtgcac ccagtcctgct caccagccc agtccgtccg gtcctcaccg 60
cctgccgggcc ggcccacccc ccaccgcagc catggacgcc atcaagaaga agatgcagat 120
gctgaaggag aacgccatcg accgcgccga gcaggccgaa gccgacaaga agcaagctga 180
ggaccgctgc aagcagctgg aggaggagca gcaggccctc cagaagaagc tgaagggggac 240
agaggatgag gtggaaaagt attctgaatc cgtgaaggag gcccaggaga aactggagca 300
ggccgagaag aaggccactg atgctgaggc agatgtggcc tcctgaacc gccgcattca 360
gctggttgag gaggagctgg accgggcccga ggagcgctg gctacagccc tgcagaagct 420
ggaggaggcc gagaaggcgg ctgatgagag cgagagagga atgaaggcca tcgaaaaccg 480
ggccatgaag gatgaggaga agatggaact gcaggagatg cagctgaagg aggccaagca 540
catcgctgag gattcagacc gcaaataatga agagggtggcc aggaagctgg tgatcctgga 600
aggagagctg gagcgctcgg aggagagggc tgagggtggcc gagagccgag ccagacagct 660
ggaggaggaa cttcgaacca tggaccaggc cctcaagtcc ctgatggcct cagaggagga 720
gtattccacc aaagaagata aatatgaaga ggagatcaaa ctgttgaggag agaagctgaa 780
ggaggetgag acccgagcag agtttgccga gaggtctgtg gcaaagtgg agaaaaccat 840
cgatgacctg gaagagacct tggccagtgc caaggaggag aacgtcgaga ttcaccagac 900
cttgagaccag accctgctgg aactcaacaa cctgtgaggg ccagccccac cccagccag 960
gctatggttg ccacccaac ccaataaaac tgatgttact agcctctcaa aaaaaaaaaa 1020

```

<210> 30

<211> 1120

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 3745193

<400> 30

```

cgggtgccacc gcggcgcaga ggagtctgca atgccgagtg gaggaaggag gaaccggagt 60
gtgagcagta gctgggtggg cagcatggct gggatcacca ccatcgaggc agtgaagcgc 120
aagatccagg ttctgcagca ccaggcagat gatgcagagg agtgagctga gcacctccag 180
tgagaagctg agggaaaaag gtgggcctgg gaacaggcag aggctgaagt ggcctccgtg 240
aacggtagga tccagctggt tgaagaggag ctggactgtg ctcaggagcg cctggccact 300
gccctgcaaa agctggaaga agcgggaaaa gctgctgatg agagtgagag agatacaaa 360
gttattgaaa tctgggcctt aaaagatgaa gaagatggaa ctccaggaaa tccaactcaa 420
agaagctaag cacattgcag atgaggcaga tgggaagtat gaagaggtgg ctcgtaagtt 480
ggtgatcatt gaaggagaca tgggatgcac agaggaacga gctgagctgg cagagtcccg 540
ttgctgagag atggatgagc agatcagact gatggaccag aacctgaagt gtctgagtgc 600
agctgaagaa aagtactctc aaaaagaaga caaatgtgag gaagagatga agattcttac 660
tgataatctc aaggaggcag agaccatgc tgagttggct gagagatcag tagccaagct 720
ggaaaagaca attgatgact tggaaagataa actgaaatgc accaaagagg aacacctctg 780
tacacaaagg atgctggacc agactttgct tgacctgaat gagatgtaga atgccccagt 840
cccaccctgc tgctgtcctt tctctgacc ctgactctgc ctgaggccag cctgcccga 900
gctgaccttt acctgagggc tgatctttaa cggaaggct gctttctcct tttgccaccc 960
cctccttccc tgctctttt tcaccaaact gtctctgcct cttcctggag attccagctg 1020
ggctagaggc tgagaacctt tgcaaacac atttaaggga atgtgagccc aatgcataat 1080
gtctttaaaa atcatgttga gaaaaaaaaa atacataata 1120

```

<210> 31  
 <211> 1471  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 3822123

<400> 31  
 caggcccagc tgagaggtgc gcgggcgagg acagcggcag cgatgcggga atgcatatca 60  
 gtccacgtgg gccaaagcggg agttcagatt ggcaatgcct gctgggagct cttctgcctg 120  
 gaacacggca tccaggcaga cggcactttt gatgctcaag ctagcaagat caacgatgat 180  
 gactccttca ccaccttttt cagcgagact ggcaatggga agcatgtgcc ccgggccgctc 240  
 atgatagatc tggagcctac tgtagtggat gaggttcggg caggaaccta ccgccagctc 300  
 ttocatccag agcagctgat cacaggaaag gaggatgcag ccaacaacta tgcccggggc 360  
 cactacacgg tgggcaagga gagcattgac ctggtgctgg accgcatacg gaagctgaca 420  
 gatgcttgct ctggcctgca gggcttctctg attttccaca gttttggtgg gggcactggc 480  
 tccggcttca cttctctgct gatggaacgc ctctccctgg attatggcaa gaaatccaag 540  
 ctggaattct ccatctaccc agccccccag gtgtctacag ccgtggtcga gccctacaac 600  
 tcttatctga ccaccacac caccctggag cactcagact gtgccttcat ggtggacaac 660  
 gaagcaatct atgacatctg ccgccgcaac ctagacatcg agcgcccaac ctacaccaac 720  
 ctcaatcgcc tcattagcca aattgtctcc tccatcacag cttctctgcg ctttgacggg 780  
 gccctcaatg tggacctgac agagtccag accaacctgg tgccctaccc tcgcatccac 840  
 ttccccctgg ccacctatgc accagtcac tctgcagaaa aggcatacca cgagcagctg 900  
 toggttggcag agatcaccaa tgcctgcttt gagcctgcca accagatggt aaagtgtgat 960  
 ccccgccacg gcaagtacat ggcctgctgc ctgctgtacc gtggagatgt ggtgccaag 1020  
 gatgtcaacg ctgccattgc cgccatcaag accaagcgca gcattcagtt tgtggactgg 1080  
 tgccccacag gcttcaaggt tggatatcaac taccagcctc cactgtggt gcctgggggt 1140  
 gacctggcca aggtgcagcg tgccgtgtgc atgctgagca acacgaccgc catcgccgag 1200  
 gcctggggcc gcctggacca caagttcgac ctgatgtatg ccaagagggc gtttgtgcac 1260  
 tggatatgtg gtgagggcat ggaggagggg gaggttctcg aggcccgta ggatatggct 1320  
 gccctggaga aggattatga ggagggtggg atcgactcct atgaggacga ggatgaggga 1380  
 gaagaataaa gcagctgcct ggagcctatt cactatgttt attgcaaaat cctttcgaaa 1440  
 taaacagttt ccttgcaaaa aaaaaaaaaa a 1471

<210> 32  
 <211> 1079  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 4217506

<400> 32  
 cagcccttct cacactcgac ccgcagaaac caccacacctt caccatgtct gacgaggaag 60  
 ttgaacaggt ggaggagcag tacgaagaag aagaggaagc ccatgaggaa gctgcagaag 120  
 tccatgagga agttcatgaa ccagaggaag ttcaagaaga caccgcagag gaggacgcgg 180  
 aagaggagaa accgagaccc aaactcactg ctcttaagat cccagaaggg gagaaagtgg 240  
 acttcgatga catccagaag aagcgtcaga acaaagacct aatggagctc caggccctca 300  
 tcgacagcca ctttgaagcc cggaagaagg aggaggagga gctggctcgt ctcaaagaga 360  
 gaatcgagaa gcgccgtgca gagagagcgg agcagcagag gattcgtgca gagaaggaga 420



gggagcgcca gaacagactg gcgaggaaaa aggccagaag ggaggaggag gatgccaaga 480  
ggagggcaga ggacgacctg aagaagaaga aagctctgtc ttccatggga gccaaactaca 540  
gcagctacct ggccaaggct gaccagaaga gaggcaagaa gcagacagcc cgggaaatga 600  
agaagaagat tctggctgag agacgcaagc cgctcaacat cgatcacctt ggtgaagaca 660  
aactgagggg caaggccaag gagctctggg agaccctgca ccagctggag attgacaagt 720  
tcgagtttgg ggagaagctg aaacgccaga aatatgacat caccacgctc aggagccgca 780  
ttgaccaggc ccagaagcac agcaagaagg ctgggacccc agccaagggc aaagtcggcg 840  
ggcgctggaa gtagagaggc cagaaaggcc cctcgaggca gagaccctcc gccctcttgc 900  
acaccagggc cgctcgtggg actccacatc ctccagcccc cacaatcctg tcaggggctc 960  
cctgacagtc ctgggggtgg agaggccatc ccggggcgtc ccccgctct gtgtccttgc 1020  
tgccttcac cctgggggc tgtgaataaa gctgcagaac ccccaaaaaa aaaaaaaaa 1079